March 1947-8

BIOMETRICS

Vol. 3/NO. 1
14

HE BIOMETRICS SECTION, AMERICAN STATISTICAL ASSOCIATION

THE ASSUMPTIONS UNDERLYING THE ANALYSIS OF VARIANCE*

CHURCHILL EISENHART

University of Wisconsin and the National Bureau of Standards

1. *Introductory Remarks.* The statistical technique known as "analysis of variance," developed more than two decades ago by R. A. Fisher to facilitate the analysis and interpretation of data from field trials and laboratory experiments in agricultural and biological research, today constitutes one of the principal research tools of the biological scientist, and its use is spreading rapidly in the social sciences, the physical sciences, and in engineering. Numerous textbooks (or, should I say "manuals"?) have been published—and, I dare say, many more are being written—that aim to provide their readers with a working knowledge of the steps of analysis-of-variance procedure with a minimum exposure to mathematical formulas and mathematical thinking. Designed expressly for the "non-mathematical reader", whose mathematical equipment is presumed to be a reasonable competence in arithmetic and elementary algebra—mere previous exposure to these subjects is not enough. The method of instruction adopted in these books consists chiefly in guiding the reader by easy stages through a series of worked examples that are typical of the more common problems amenable to analysis of variance that arise in the scientific or engineering field with which the author of the book concerned is conversant.[1]

* An expository address delivered at a joint session of the Biometrics Section of the American Statistical Association and the Institute of Mathematical Statistics, held on December 28, 1946, in conjunction with the 113th Annual Meeting of the American Association for the Advancement of Science, Boston, Massachusetts.

[1] The author has found the discussions and examples of analysis-of-variance procedures given in the following four books especially valuable both for reference and for purposes of instruction: C. H. Goulden, *Methods of Statistical Analysis;* G. W. Snedecor, *Calculation and Interpretation of Analysis of Variance and Covariance;* G. W. Snedecor, *Statistical Methods;* L. H. C. Tippett, *The Methods of Statistics.* See full bibliographical references at the end of this paper.

These introductions to analysis of variance have been definitely worthwhile in at least three respects: first, they have acquainted a larger audience with the procedures of analysis of variance and its value as a research tool than probably would have been achieved by more mathematical expositions of the subject (even unfavorable reviews of some of these books have focused attention on the analysis of variance itself as a research tool "that needs further looking into"); second, by studying the worked examples provided and by carrying through analogous steps with data of their own, readers of these books have developed an amazing proficiency with the arithmetical steps involved, even in the cases of analyses associated with fairly complicated experimental designs which probably would not have been attempted or, if attempted, almost certainly would not have been analyzed correctly without the aid of these books; and third, since the worked examples given in these books have generally illustrated statistically sound experimental designs which were more efficient than the designs previously used by their readers, these readers have frequently adopted analogous designs in their own research (in order to be able to follow the book when the data are in and are crying for analysis), with a resulting general improvement of research procedure.

The principal deficiency of these books has been their failure to state explicitly the several asumptions underlying the analysis of variance, and to indicate the importance of each from a practical viewpoint. The mathematical treatments of analysis of variance have shared this deficiency to some extent, for, while they have posited the necessary and sufficient conditions[2] for strict validity of the entire set of analysis-of-variance procedures and associated tests of significance, they have not generally indicated in sufficient detail the actual functions of the respective assumptions—1) which can be dispensed with for certain purposes; 2) which are absolutely necessary, and what are likely to be the consequences if these are not fulfilled; and 3) what can be done "to bring into line," for purposes of analysis, data which in their original form are not amenable to analysis of variance.[3] In this paper I shall go into these matters in some detail. My assignment is to

[2] The conditions here referred to are certainly sufficient: they may be necessary in the mathematical sense, but no proof of this is known to the writer. For a variety of reasons he believes them to be "necessary in practice" in the same sense that, if there are exceptions, the circumstances required would be regarded by the practical man as "pathological."

[3] See, for example, the discussions of analysis of variance in H. Cramér, *Mathematical Methods of Statistics;* in M. G. Kendall, *The Advanced Theory of Statistics*, Volume II; and in S. S. Wilks, *Mathematical Statistics*. Somewhat more complete discussions have been given by J. O. Irwin in his paper entitled "Mathematical Theorems Involved in the Analysis of Variance" and in his note "On the Independence of the Constituent Items in the Analysis of Variance."

enumerate the several assumptions underlying the analysis of variance and to point out the practical importance of each. As we shall see, these assumptions are quite simple to state, and the practical significance of each not difficult to grasp. Professor Cochran, in the second paper of this issue of *Biometrics,* tells of some of the consequences to expect when certain of these assumptions are not fulfilled. Finally, Professor Bartlett, in the third paper, indicates how, by the use of transformations, some of these consequences can be avoided and valid conclusions reached by analysis of variance, when the data in their original form are essentially intractable by analysis of variance.

2. *Two Distinct Classes of Problems Solvable by Analysis of Variance.* Turning now to my assignment, I am obliged at the outset to draw attention to the fact that analysis of variance can be, and is, used to provide solutions to problems of two fundamentally different types. These two distinct classes of problems are:

(2.1) *Class I: Detection and Estimation of Fixed (Constant) Relations Among the Means of Sub-Sets of the Universe of Objects Concerned.* This class includes all of the usual problems of estimating, and testing to determine whether to infer the existence of, *true* differences among "treatment" means, among "variety" means, and, under certain conditions, among "place" means. Included in this class are all the problems of univariate and multivariate regression and of harmonic analysis. With respect to the problems of estimation belonging to this class, analysis of variance is simply a form of the method of least squares: the analysis-of-variance solutions are the least-squares solutions. The cardinal contribution of analysis of variance to the actual procedure is the *analysis-of-variance table* devised by R. A. Fisher, which serves to simplify the arithmetical steps and to bring out more clearly the significance of the results obtained. The analysis-of-variance tests of significance employed in connection with problems of this class are simply extensions to small samples of the theory of

Entered as second-class matter, May 25, 1945, at the post office at Washington, D. C., under the Act of March 3, 1879. *Biometrics* is published four times a year—in March, June, September and December—by the American Statistical Association for its Biometrics Section. Editorial Office: 1603 K Street, N. W., Washington 6, D. C.

Membership dues in the American Statistical Association are $5.00 a year, of which $3.00 is for a year's subscription to the quarterly *Journal,* fifty cents is for a year's subscription to the *ASA Bulletin.* Dues for Associate members of the Biometrics Section are $2.00 a year, of which $1.00 is for a year's subscription to *Biometrics.* Single copies of *Biometrics* are $1.00 each and annual subscriptions are $2.00. Subscriptions and applications for membership should be sent to the American Statistical Association, 1603 K Street, N. W., Washington 6, D. C.

least squares developed by Gauss and others—the extension of the theory to small samples being due principally to R. A. Fisher.

(2.2) *Class II: Detection and Estimation of Components of (Random) Variation Associated with a Composite Population.* This class includes all problems of estimating, and testing to determine whether to infer the existence of, components of variance ascribable to random deviation of the characteristics of individuals of a particular generic type from the mean values of these characteristics in the ''population'' of all individuals of that generic type, etc. In a sense, *this is the true analysis of variance,* and the estimation of the respective components of the over-all variance of a single observation requires further steps beyond the evaluations of the entries of the analysis-of-variance table itself. Problems of this class have received considerably less attention in the literature of analysis of variance than have problems of Class I.[4]

The failure of most of the literature on analysis of variance to focus attention on the distinction between problems of Class I and problems of Class II is very likely due to two facts: first, the literature of analysis of variance deals largely with tests of significance in contrast to problems of estimation; second, when analysis of variance is used merely to determine whether to infer (a) the existence of fixed differences among the true means of the sub-sets concerned or (b) the existence of a component of variance ascribable to a particular factor, the computational procedure and the mechanics of the statistical tests of significance are the same in either case—the same test criterion (*F or z*) is evaluated and referred to the same ''levels of significance'' in either case. On the other hand, in the estimation of the relevant parameters, and in the evaluation of the efficiency or resolving power of a particular experimental design, the distinction between these two classes of problems needs to be taken into account, since in problems of Class I the parameters involved are *means* and the issues of interest are concerned with the interrelations of these means, i.e., with the differences between pairs of them, with their functional dependence on some independent variable(s), etc.; whereas in problems of Class II the parameters involved are *variances* and their absolute and relative magnitudes are of primary importance. In other words, the mathematical models appropriate to problems of Class I differ from the mathematical models

[4] R. A. Fisher gives a brief discussion of estimating, and testing for the existence of components of variance in Section 40 of his *Statistical Methods for Research Workers.* Tippett considers such problems in Sections 6.1–6.2, 6.4, 10.11, and 10.3. Snedecor's treatment is somewhat more complete: *Statistical Methods,* Sections 10.6–10.12, 11.4, 11.7–11.8, 11.14, 11.16. The most complete discussions of problems of Class II are H. E. Daniels, "The Estimation of Components of Variance," and S. Lee Crump, "The Estimation of Variance Components in Analysis of Variance."

appropriate to problems of Class II, and consequently, so do the questions to be answered by the data.

3. *The Algebra of Analysis of Variance.* It was remarked above that the computational steps leading to an analysis-of-variance table are the same for problems of Class I and Class II. This is due largely to the fact that the decomposition of the (total) sum of squared deviations of the individual observations from the general mean of the observations into two or more "sums of squares" is based in every case upon an algebraic identity (appropriate to the case concerned) that is valid whatever the meanings of the numbers involved. To demonstrate this in complete generality would render the substance of this paper somewhat complicated, and these complexities would, I fear, distract attention from the main theme. Accordingly, I shall restrict myself to consideration of the algebra of the decomposition for the case of rc numbers arranged in a rectangular array of r rows and c columns. In order to be able to identify the various numbers, let us denote by x_{ij} the number occurring in the i^{th} row and j^{th} column of this array. If we border the rectangular array with a column of row means and a

TABLE 1

Column

Row	1	2	3	·	·	·	·	·	j	·	·	c	Row Means
1	X_{11}	X_{12}	X_{13}	·	·	·	·	·	X_{1j}	·	·	X_{1c}	$X_{1.}$
2	X_{21}	X_{22}	X_{23}	·	·	·	·	·	X_{2j}	·	·	X_{2c}	$X_{2.}$
·	·	·	·	·	·	·	·	·	·	·	·	·	·
·	·	·	·	·	·	·	·	·	·	·	·	·	·
·	·	·	·	·	·	·	·	·	·	·	·	·	·
i	X_{i1}	X_{i2}	X_{i3}	·	·	·	·	·	X_{ij}	·	·	X_{ic}	$X_{i.}$
·	·	·	·	·	·	·	·	·	·	·	·	·	·
·	·	·	·	·	·	·	·	·	·	·	·	·	·
r	X_{r1}	X_{r2}	X_{r3}	·	·	·	·	·	X_{rj}	·	·	X_{rc}	$X_{r.}$
Col. Means	$X_{.1}$	$X_{.2}$	$X_{.3}$	·	·	·	·	·	$X_{.j}$	·	·	$X_{.c}$	$X_{..}$

row of column means, then we have a situation such as that portrayed in Table 1, where $x_{i.}$ denotes the arithmetic mean of the c values of x in the i^{th} row, $x_{.j}$ denotes the arithmetic mean of the r values of x in the j^{th} column, and $x_{..}$ denotes the arithmetic mean of all the rc values in the array.

It is evident that the following is an algebraic identity *whatever the interpretation of the numbers* x_{ij} *involved:*

$$(1) \qquad (x_{ij}-x_{..}) = (x_{i.}-x_{..}) + (x_{.j}-x_{..}) + (x_{ij}-x_{i.}-x_{.j}+x_{..}).$$

Remembering that by definition the arithmetic mean of m values of a quantity y is (sum of the m values)$/m$ we see that

$$(2) \qquad \text{Arithmetic Mean of } y \equiv \bar{y} \equiv \frac{1}{m} S(y) \text{ implies } S(y-\bar{y}) = 0,$$

and

$$(3) \qquad S(y-\bar{y})^2 = S(y^2) - \frac{[S(y)]^2}{m},$$

where S denotes summation over all the values of y involved. Squaring both sides of (1) and summing over all rc observations, the algebraic identity

$$(4) \qquad \underset{(A)}{S(x_{ij}-x_{..})^2} = \underset{(B)}{S(x_{i.}-x_{..})^2} + \underset{(C)}{S(x_{.j}-x_{..})^2} + \underset{(D)}{S(x_{ij}-x_{i.}-x_{.j}+x_{..})^2}$$

results, where S denotes summation over all the values in the entire array; the cross-products involved sum to zero by virtue of (2), *on account of the fact that* $x_{i.}$, $x_{.j}$, *etc., and* $x_{..}$ *are means.* The (A), (B), (C), and (D) sums of squared quantities in (4) are what are usually referred to in an analysis-of-variance table as the "total", the "between-row-means", the "between-column-means", and "residual" sums of squares, respectively. Since $(x_{i.}-x_{..})^2$ is identically the same for each of the c observations in the i^{th} row, and $(x_{.j}-x_{..})^2$ is the same for each observation in the j^{th} column, it is sometimes convenient to write (4) as

$$(5) \qquad \underset{(A)}{\sum_{i=1}^{r}\sum_{j=1}^{c}(x_{ij}-x_{..})^2} = \underset{(B)}{c\sum_{i=1}^{r}(x_{i.}-x_{..})^2} + \underset{(C)}{r\sum_{j=1}^{c}(x_{.j}-x_{..})^2} + \underset{(D)}{\sum_{i=1}^{r}\sum_{j=1}^{c}(x_{ij}-x_{i.}-x_{.j}+x_{..})^2}$$

where $\sum\limits_{i=1}^{r}\sum\limits_{j=1}^{c}$ denotes summation over all the observations in the array, $\sum\limits_{i=1}^{r}$ denotes summation only over i, $i=1$ to $i=r$, and $\sum\limits_{j=1}^{c}$ denotes summation only over j, for $j=1$ to $j=c$.

(3.1) *"Practical" Formulas.* With the aid of the identity (3), it is easy to derive the "practical" formulas used for calculation:

$$\text{(A)}\quad \sum_{i=1}^{r}\sum_{j=1}^{c}(x_{ij}-x_{..})^2=\sum_{i=1}^{r}\sum_{j=1}^{c}(x_{ij}{}^2)-\frac{[\Sigma\Sigma(x_{ij})]^2}{rc}=$$

$$\text{Sum of the squares of all observations}-\frac{[\text{Sum of all observations}]^2}{\text{Number of observations}},$$

$$\text{(B)}\quad c\sum_{i=1}^{r}(x_{i.}-x_{..})^2=\sum_{i=1}^{r}\frac{(cx_{1.})^2}{c}-\frac{\sum\limits_{i=1}^{r}(cx_{i.})^2}{cr}$$

$$\text{(6)}\quad =\text{Sum with respect to } i \text{ of } \frac{(i^{\text{th}}\text{-row Total})^2}{c}$$

$$-\{\text{Correction Term, given above}\},$$

$$\text{(C)}\quad r\sum_{j=1}^{c}(x_{.j}-x_{..})^2=\sum_{j=1}^{c}\frac{(rx_{.j})^2}{r}-\frac{\left[\sum\limits_{j=1}^{c}(rx_{.j})\right]^2}{rc}=$$

$$\text{Sum with respect to } j \text{ of } \frac{(j^{\text{th}}\text{-Column Total})^2}{r}-\textit{idem},$$

$$\text{(D)}\quad \sum_{i=1}^{r}\sum_{j=1}^{c}(x_{ij}-x_{i.}-x_{.j}+x_{..})^2=\text{(A)}-\text{(B)}-\text{(C)}.$$

I repeat: All of the familiar formulas and procedures for evaluating component "sums of squares" that add up to the "total sum of squares" are based on algebraic identities, and are valid as descriptions of properties of the data whatever the interpretation of the numbers involved.. Indeed, the fact that the "components" add up to the total is an algebraic (or, should I say a "geometric") property and means that (and will only happen when) the respective component "sums of squares" are themselves the squares of, or sums of the squares of, linear combinations of the observations that summarize mutually distinct properties of the data, or, as a geometer would say, linear combinations that define mutually orthogonal vectors in the N-dimensional sample space.[5] Similarly, all of the familiar formulas and procedures for evaluating regression coefficients and the sum of squared deviations from the fitted regression, when the fitting is by the method of least squares, are based upon algebra and calculus, and the results obtained are valid as descriptions of properties of the data in hand, whatever the interpretation of the numbers involved.

In summary, when the formulas and procedures of analysis of variance are used merely to summarize properties of the data in hand, no assumptions are needed to validate them. On the other hand, when analysis of variance is used as a method of statistical inference, for inferring properties of the "population" from which the data in hand were drawn, then certain assumptions, about the "population" and the sampling procedure by means of which the data were obtained, must be fulfilled if the inferences are to be valid.

4. *The Assumptions Underlying the Use of Analysis of Variance as a Method of Statistical Inference.* As was remarked earlier, analysis of variance can be, and is, used to provide solutions to two fundamentally different types of problems: On the one hand, it can be used to detect the existence of, and to estimate the parameters defining, fixed (constant) relations among the population means. These were referred to as problems of Class I. On the other hand, analysis of variance can be used to detect the existence of, and to estimate, components of variance. These were termed problems of Class II. To formulate with complete generality the mathematical models upon which the solutions of problems of Class I and Class II by analysis of variance are based would render the substance of this paper somewhat complicated from this point on, and would, I fear, divert attention from the really important distinctions between the two different models, and from the differences between the assumptions required in order to be able to draw valid inferences by analysis of variance in the two cases. Therefore,

[5] To see what we mean by "mutually distinct" and by "orthogonal" in practical language, let us note that, if in the case of numbers arranged as in Table 1 we add a single arbitrary constant to each of the numbers in the first column, a different arbitrary constant to each of the numbers in the second column, and so forth through the c^{th} column, then the several row means will be altered by different amounts, which will be determined by the actual constants added, but the several row means will all be altered by the same amount, so that the *difference* between any pair of row means, $(x_{i_1\cdot} - x_{i_2\cdot})$, will be unchanged. Similarly the values of such quantities as $(x_{i\cdot} - x_{\cdot\cdot})$ and $(x_{ij} - x_{i\cdot} - x_{\cdot j} + x_{\cdot\cdot})$ will be unchanged by this tampering with the columns, so that the "between-row-means" and the "residual" sums of squares will be unchanged also. This is because *differences among row means* (or *differences of row means from the general mean*) and the *residuals* are *orthogonal* to *differences among column means* (or *differences of column means from the general mean*), that is, summarize mutually distinct properties of the actual numbers involved. This little trick of adding arbitrary constants in accordance with a definite pattern is a convenient practical way of checking whether particular combinations of the observations are mutually orthogonal.

two different models appropriate to data arranged as in Table 1 will be considered in detail and the relation of each assumption to the inferential steps indicated:

(4.1) *Model I, Special Case: Parameters Are Population Means.* Numbers x_{ij} arranged as in Table 1 do not lie within the province of mathematical statistics, nor can any statistical inferences be based upon them, unless it is assumed that they are (observed values of) *random variables* of some sort. Therefore, in order to bring the discussion within the province of statistical inference we must make

> *Assumption 1* (*Random Variables*): The numbers x_{ij} are (observed values of) random variables that are distributed about true mean values m_{ij}, $(i = 1, 2, \ldots, r; j = 1, 2, \ldots, c)$, that are fixed *constants.*

In statistical language this assumption states that, if some particular type of experiment leading to numbers arranged as in Table 1 were repeated indefinitely, then the numbers occurring in the i^{th} cell of the j^{th} column would vary at random about an average value equal to m_{ij}, which is, therefore, a parameter that characterizes the expected value of the number x_{ij}. If, for example, the several rows of Table 1 correspond to different "varieties" and the several columns to different "treatments," then m_{ij} is the so-called *true* (or expected) *yield* of the i^{th} "variety" when subjected to the j^{th} "treatment," under certain growing conditions.

Clearly the parameters m_{ij} can be arranged in a table analogous to Table 1, and bordered by the row-wise means $m_{i.}$, $(i = 1, 2, \ldots, r)$, and the column-wise means $m_{.j}$, $(j = 1, 2, \ldots, c)$ of these parameters, to which may be added, in the lower right corner, the mean $m_{..}$ of all rc of these parameters. If one is merely interested in obtaining unbiased estimates of mean differences such as $m_{12} - m_{52}$, e.g., of the mean difference between variety 1 and variety 5 under treatment 2, then Assumption 1 is sufficient, and $x_{12} - x_{52}$ provides the desired estimate. More generally, Assumption 1 implies that an unbiased estimator of any linear function of the m_{ij} with *known* coefficients is provided by the same linear function of the x_{ij}. Furthermore, if the variances of the x_{ij} about their respective means and their intercorrelations are known, then the variance of any linear function of the x_{ij} can be evaluated, and provides a measure of the precision of this linear function of the x_{ij} as an unbiased estimator of the corresponding linear function of the m_{ij}.

On the other hand, when the entries m_{ij} of such a table of *true* *means* are simple *additive functions* of the corresponding marginal means and the general mean, that is, when

$$m_{ij} = m_{..} + (m_{i.} - m_{..}) + (m_{.j} - m_{..}), \tag{7}$$

for $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, c$, then the statistical inferences that may be based upon the x_{ij} are of a much more satisfactory sort. For instance, when (7) is satisfied, the difference between an arbitrary pair of row-wise marginal means, e.g., $m_{1.}$ and $m_{2.}$, is a *comprehensive* measure of the average difference in effectiveness of the factors identified with these rows. When (7) is not satisfied, then $m_{1.} - m_{2.}$ is merely a measure of the average difference between the effects of the corresponding row factors *when the column factors are as in the experiment concerned.* In other words, when additivity, as defined by (7), does not obtain, then it is not possible to define *the* mean difference in effectiveness of any given pair of the row factors, since the actual mean difference in effectiveness of these row factors will depend upon the column factor(s) concerned; and, conversely, the actual mean difference in effectiveness of a pair of column factors will depend upon the row factor(s) concerned. Hence, when additivity does not prevail, we say that there are *interactions* between row factors and column factors. Thus, in the case of varieties and treatments considered above, additivity implies that, under the general experimental conditions of the test, the true mean yield of one variety is greater (or, less) than the true mean yield of another variety by an amount—an additive constant, not a multiplier—that is the same for each of the treatments concerned, and, conversely, the true mean yield with one treatment is greater (or, less) than the true mean yield with another treatment by an amount that does not depend upon the variety concerned; which is exactly what is meant when we say that there are no "interactions" between varietal and treatment effects.

Therefore, in order to dispense with interactions and thus make possible the drawing of general inferences from the x_{ij}, let us make

Assumption 2 (*Additivity*) :[6] the parameters m_{ij} are related to the

[6] In its most general form Model I involves N random variables $x_1, x_2, \ldots, x_N$ with mean values m_i, $(i = 1, 2, \ldots, N)$, and it is assumed that the m_i are linear functions of $s < N$ *unknown* parameters θ_j, $(j = 1, 2, \ldots, s)$, with *known* coefficients, c_{ij}, the matrix of which is non-singular; thus

$$m_i = c_{i1}\theta_1 + c_{i2}\theta_2 + \ldots + c_{is}\theta_s \qquad (i = 1, 2, \ldots, N),$$

and non-singularity of the matrix $\| c_{ij} \|$ signifies that from this set of N equations it is possible to select at least one system of s equations that is soluble with respect to the θ's.

This is known as the *general linear hypothesis*. For details, see the papers by F. N. David and J. Neyman, by S. Kolodziejczyk, and by P. C. Tang cited in the list of references.

means $m_{i.}$, $m_{.j}$, and $m_{..}$ as specified by (7), for $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, c$.

When *Assumption 1* and *Assumption 2* are satisfied, then the difference between any pair of row-wise means of the observations x_{ij}, e.g., $x_{2.} - x_{5.}$, is an unbiased estimator of the *general* average difference in effectiveness of the row factors concerned, i.e., of $m_{2.} - m_{5.}$ in this case; and, similarly, the difference between any pair of column-wise means of the observations is an unbiased estimator of the *general* average difference in effectiveness of the column factors concerned. Furthermore, since such estimators are linear functions of the x_{ij}, the variances of these estimators can be evaluated readily when the variances and intercorrelations of the x_{ij} are known. On the other hand, if these variances and intercorrelations are unknown—the usual case in practice—then it is not possible to derive from the observed values of the x_{ij}, $(i = 1, 2, \ldots, r; j = 1, 2, \ldots, c)$, an unbiased estimate of the variance of any single x_{ij}, or of any particular linear combination of them, unless certain additional conditions are fulfilled by the x_{ij}. For instance, if the x_{ij} are mutually uncorrelated,[7] and if variance of the x_{ij} are given by

$$\text{(8)} \qquad \text{variance of } x_{ij} = \frac{\sigma^2}{w_{ij}}$$

where the relative "weights," w_{ij}, are *known* constants, $(i = 1, 2, \ldots, r; j = 1, 2, \ldots, c)$, and σ^2 is an *unknown* constant, then an unbiased estimate of σ^2, and thence unbiased estimates of the variances of linear combinations of the x_{ij}, can be derived from the observations x_{ij} by the method of least squares. For details see, for example, the paper by F. N. David and J. Neyman cited in the list of references. They assume the x's to be mutually independent, whereas it is sufficient to assume that they are mutually uncorrelated.

It should be noted here that thus far the only motivation that has been given for the making of *Assumption 2* is the more general nature of the inferences that may be drawn from the observed means $x_{i.}$ and $x_{.j}$, $(i = 1, 2, \ldots, r; j = 1, 2, \ldots, c)$, when it is satisfied. We shall now show that, in general, it is not possible to derive from the observations x_{ij} by the *usual analysis-of-variance procedures,* unbiased estimates of the variances of the x_{ij}, and thence of any particular linear combinations of them, unless *Assumption 1, Assumption 2,* and *Assumption 3,* given below, are *all* satisfied.

[7] That is, if the *covariances* $\mathcal{E}\{[x_{ij} - \mathcal{E}(x_{ij})]\ [x_{pq} - \mathcal{E}(x_{pq})]\}$, where $(i, j) \neq (p, q)$, (i and $p = 1, 2, \ldots, r$; j and $q = 1, 2, \ldots, c$), and $\mathcal{E}$ denotes "expected value of," are *all* equal to zero.

Assumption 3 (*Equal Variances and Zero Correlations*): The random variables x_{ij} are *homoscedastic* and *mutually uncorrelated,* that is, they have a common variance σ^2 and all covariances among them are zero.

The foregoing pronouncement can be demonstrated readily by considering the analysis-of-variance table shown as Table 2. This represents the situation when *Assumption 1* and *Assumption 3* are both satisfied, but *Assumption 2* is not. We notice that under these conditions each of the "mean squares" customarily evaluated in such cases will have, in general, an expected value larger than σ^2. If, on the other hand, *Assumption 2* is satisfied also, then the "residual" mean square will be an unbiased estimator of σ^2, the variance of any single observation x_{ij}. This situation is portrayed in Table 3. Hence, when *Assumption 1, Assumption 2,* and *Assumption 3* are all satisfied, an unbiased estimate of the variance of the difference of two *observed* row means can be evaluated from 2(residual mean square)/*c;* and an unbiased estimate of the variance of the difference of two *observed* column means, from 2(residual mean square)/*r*. Furthermore, under these conditions the between-row-means mean square in general will *tend* to exceed the residual mean square, and this tendency will be greater when the true row means, the $m_{i.}$, differ markedly in magnitude than when they differ only slightly. Similarly, the between-column-means mean square in general will tend to exceed the residual mean square by an amount that depends upon the degree of "scatter" of the true column means, the $m_{.j}$, about $m_{..}$, the mean of all the m_{ij}. Thus we have yardsticks for

TABLE 2

ANALYSIS OF VARIANCE

(Non-Additive Case)

Variation	Degree of Freedom	Sums of Squares	Mean Square	Expected Value of Mean Square
Between Row Means	$r-1$	$S(X_{i.}-X_{..})^2$	$S(X_{i.}-X_{..})^2/(r-1)$	$\sigma^2+S(m_{i.}-m_{..})^2/(r-1)$
Between Column Means	$c-1$	$S(X_{.j}-X_{..})^2$	$S(X_{.j}-X_{..})^2/(c-1)$	$\sigma^2+S(m_{.j}-m_{..})^2/(c-1)$
Residual	$(r-1)(c-1)$	$S(X_{ij}-X_{i.}-X_{.j}+X_{..})^2$	$S(X_{ij}-X_{i.}-X_{.j}+X_{..})^2/(r-1)(c-1)$	$\sigma^2+S(m_{ij}-m_{i.}-m_{.j}+m_{..})^2/(r-1)(c-1)$
Total	$rc-1$	$S(X_{ij}-X_{..})^2$	$S(X_{ij}-X_{..})^2/(rc-1)$	$\sigma^2+S(m_{ij}-m_{..})^2/(rc-1)$

judging whether there exist real differences among the true means for the row factors, and for the column factors. Unfortunately, however, our yardsticks have no scales, i.e., probability levels, marked on them, so that with them we cannot conduct exact tests of significance corresponding to previously agreed upon probability levels. In order to be able to do this, the form of the joint distribution of the x_{ij} must be specified. To this we shall return in a moment.

TABLE 3

ANALYSIS OF VARIANCE
(Additive Case)

Variation	Degree of Freedom	Sum of Squares	Mean Square	Expected Mean Square
Between Row (Variety) Means	$r-1$	$S(X_{i.}-X_{..})^2$	$S(X_{i.}-X_{..})^2/(r-1)$	$\sigma^2+S(m_{i.}-m_{..})^2/(r-1)$
Between Column (Cultivation) Means	$c-1$	$S(X_{.j}-X_{..})^2$	$S(X_{.j}-X_{..})^2/(c-1)$	$\sigma^2+S(m_{.j}-m_{..})^2/(c-1)$
Residual	$(r-1)(c-1)$	$S(X_{ij}-X_{i.}-X_{.j}+X_{..})^2$	$S(X_{ij}-X_{i.}-X_{.j}+X_{..})^2/(r-1)(c-1)$	σ^2
Total	$rc-1$	$S(X_{ij}-X_{..})^2$	$S(X_{ij}-X_{..})^2/(rc-1)$	$\sigma^2+\dfrac{S(m_{i.}-m_{..})^2+S(m_{.j}-m_{..})^2}{rc-1}$

At this juncture let us pause for an instant to note that *it has not been necessary to postulate mutual* INDEPENDENCE *of the x_{ij} in order to achieve Table 3 and the results deducible from it*—for these, *existence of the mean values of the x_{ij}* (Assumption 1), *additivity* (Assumption 2), *and equal variances and zero covariances* (Assumption 3) *are sufficient.*

Also, let us examine the situation where *Assumption 1* and *Assumption 2* are satisfied, but *Assumption 3* is not. In this case the four values of σ^2 that appear in the last column of Table 3 must be replaced, in general, by four *different* quantities, which we may denote by σ_1^2, σ_2^2, σ_3^2, and σ_4^2. In general these will be complex weighted means of the variances and covariances of the x_{ij}, and the neatness of Table 3 is lost.

In summary, if *Assumption 1* is satisfied, but if either *Assumption 2,* or *Assumption 3,* or both, is (are) not satisfied, then the strict validity of analysis of variance as a method of solution of problems of Class I vanishes out the window.

Finally, even when *Assumption 1, Assumption 2,* and *Assumption 3* are satisfied, it is still not possible to conduct exact tests of significance based on the x_{ij} alone, e.g., tests of significance based upon Fisher's z- or Snedecor's F-distributions. Fortunately, *normality,* in addition to *Assumptions 1–3,* is sufficient for exact tests of significance. Therefore let us make

> *Assumption 4* (*Normality*): The x_{ij} are jointly distributed in a multivariate normal (Gaussian) distribution.

It may be noted that when *Assumption 4* is satisfied, *Assumption 1* is partially redundant, and serves principally to define the parameters m_{ij}. Furthermore, zero covariances, as postulated in *Assumption 3,* taken in conjunction with normality, postulated in *Assumption 4,* imply mutual independence of the x_{ij}. Thus independence finally sneaks in by the back door, so to speak.

When *Assumptions 1–4* are all satisfied, then all of the usual analysis-of-variance procedures for estimating, and testing to determine whether to infer the existence of, *fixed linear relations,* e.g., non-zero differences, among population *means,* are strictly valid. In particular, an unbiased estimator of any given linear function of the parameters m_{ij} is provided by the identical linear function of the observations x_{ij}, an unbiased estimate of its variance can be derived from the "residual" mean square and exact confidence limits for the value of the given linear function of the parameters can be deduced with the aid of Student's t-distribution. Furthermore, when the row-wise population means, the $m_{i.}$, are all equal, then the quotient ("between-row-means" mean square)/("residual" mean square) will be distributed according to Snedecor's F-distribution for $n_1 = (r-1)$ and $n_2 = (r-1)(c-1)$ degrees of freedom, respectively, which is the basis of the customary *test* of the hypothesis that the $m_{i.}$ are all equal, and the *power* of the test can be evaluated from the tables provided by P. C. Tang, and by Emma Lehmer—see references. An analogous statement can be made with respect to the column-wise population means, the $m_{.j}$.

Therefore, we can summarize the foregoing by the following theorem:

> THEOREM I: *The necessary*[8] *and sufficient conditions for the strict validity of analysis-of-variance procedures for solving problems of Class I with respect to data arranged as in Table 1 are that*

$$x_{ij} = m_{..} + (m_{i.} - m_{..}) + (m_{.j} + m_{..}) + z_{ij}, \qquad (i = 1, 2, \ldots, r;\ j = 1, 2, \ldots, c) \tag{9}$$

[8] See footnote 2.

where the $m_{i.}$, $m_{.j}$, and $m_{..}$ are constants with

$$(10) \qquad m_{..} = \sum_{i=1}^{r} m_{i.}/r = \sum_{j=1}^{c} m_{.j}/c$$

and the z_{ij} are normally and independently distributed about zero with a common variance σ^2.

(4.2) *Model II: Parameters are Components of Variance.* The preceding discussion of the application of analysis of variance as a method of drawing statistical inferences about the parameters involved in the mathematical model of an experiment leading to numbers arranged as in Table 1 has been concerned entirely with the problems of Class I, where the parameters are *means* and the object of the analysis is to estimate these means or to infer whether certain differences among them are or are not zero. We shall now consider the application of analysis of variance as a method of statistical inference with respect to *components of variance* involved in the mathematical model of an experiment leading to numbers arranged as in Table 1.

For the sake of concreteness, let us suppose for the moment that r animals are drawn at random from the available (large) stock of a given species and that some characteristic of each, say its body temperature, is measured on each of c days randomly located throughout some period of time. Such measurements could be arranged as in Table 1. Furthermore, let us suppose that our ultimate objective is to determine very precisely *the* body temperature characteristic of this species. By *the* body temperature characteristic of this species we mean that value about which the body temperatures of individual animals from the species will vary as a result of *biological variation*, this variability being accentuated, possibly, by day-to-day vicissitudes in the case of each animal. Under these circumstances it will clearly be of interest (a) to ascertain whether there is a component of variation assignable to day-to-day changes in the body temperature of a single animal, and (b) to compare its magnitude with the component of variation assignable to animal-to-animal variability within the species, in order to have a basis for deciding whether in collecting further data a few animals examined on each one of many days, or many animals examined on each one of only a few days, will lead to a more precise estimate of the mean body temperature characteristic of the species.

These questions may be answered by analysis of variance by arranging the data as in Table 1 and making the following assumptions:

Assumption A (*Random Variables*) : The numbers x_{ij} are (observed values of) random variables that are distributed about a common mean value $m_{..}$, $(i=1, 2, \ldots, r; j=1, 2, \ldots, c)$, where $m_{..}$ is some fixed constant.

Assumption B (*Additivity of Components*): The random variables x_{ij} are sums of component random variables, thus

$$(11) \quad x_{ij} = m_{..} + (m_{i.} - m_{..}) + (m_{.j} - m_{..}) + z_{ij}, \qquad (i=1, 2, \ldots, r; j=1, 2, \ldots, c)$$

where the $(m_{i.} - m_{..})$, the $(m_{.j} - m_{..})$, and the z_{ij} are random variables.[9]

It should be noted that *Assumption A* in conjunction with *Assumption B* implies that the mean values of the $(m_{i.} - m_{..})$, of the $(m_{.j} - m_{..})$ and of the z_{ij}, are all zero.

Assumption C (*Zero Correlations and Homogeneous Variances*): The random variables $(m_{i.} - m_{..})$, $(m_{.j} - m_{..})$, and z_{ij} are distributed with variances σ_r^2, σ_c^2, and σ^2, respectively, and all covariances among them are zero.

By following a line of reasoning similar to that presented in detail in the preceding section for the case of Model I, it is clear that here, in the case of Model II, the principal function of *Assumption A* is to bring the problem within the province of mathematical statistics; of *Assumption B*, to give specific meaning to the concept of "components of variance"; and of *Assumption C*, to dispense with interactions and render each of the "components of variance" assignable to a distinct "factor." It should be noted, however, that *independence* of the respective component deviations $(m_{i.} - m_{..}, m_{.j} - m_{..}$, and $z_{ij})$ of an x_{ij} from the general population mean $(m_{..})$ is not assumed—it is merely assumed that all covariances among them are zero, i.e., that they are *mutually* uncorrelated.

Collectively, *Assumptions A, B, and C* imply that

(12)

$$\sigma^2_{xij} \equiv \text{variance of a single observation} \equiv \mathcal{E}\,(x_{ij} - m_{..})^2 = \sigma^2 + \sigma_r^2 + \sigma_c^2$$

$$\sigma^2_{xi\,.} \equiv \text{variance of a row-wise mean} \equiv \mathcal{E}\,(x_{i.} - m_{..})^2 = \sigma_r^2 + \frac{\sigma^2}{c}$$

$$\sigma^2_{x\,.\,j} \equiv \text{variance of a column-wise mean} \equiv \mathcal{E}\,(x_{.j} - m_{..})^2 = \sigma_c^2 + \frac{\sigma^2}{r}$$

[9] In the example considered above, $(m_{i.} - m_{..})$ represents the deviation of the long-run mean body temperature of the i^{th} animal from the long-run mean body temperature of the species; similarly, $(m_{.j} - m_{..})$ is an adjustment for the jth day, assumed applicable to the body temperature of any animal from the species on that day. The z_{ij} are "catch-alls" and represent errors of measurement, etc.

$\sigma_{\bar{x}}^2 . . \equiv$ variance of the general mean $\equiv \mathcal{E}\ (x_{..} - m_{..})^2 = \dfrac{\sigma_r{}^2}{r} + \dfrac{\sigma_c{}^2}{c} + \dfrac{\sigma^2}{rc}$

Whence the expected values of the several mean squares of the customary analysis-of-variance table are as shown in the last column of Table 4. In brief, when *Assumptions A, B, and C* are satisfied, the residual mean square is an unbiased estimate of the "residual" variance, σ^2; subtracting the residual mean square from the between-row-means mean square and dividing this difference by c, the number of columns, yields an unbiased estimate of the "row-factor" component of variance, $\sigma_r{}^2$. By a similar procedure an unbiased estimate of "column-factor" component of variance, $\sigma_c{}^2$, can be obtained. It may be noted in passing that *the "naive" estimate of the over-all variance of a single observation, furnished by the "total" mean square, is a biased estimate,* and becomes unbiased only asymptotically as both r and c increase indefinitely.

TABLE 4

ANALYSIS OF VARIANCE

(Additive Case, Row- and Column-Factors *Random*)

Variation	Degree of Freedom	Sum of Squares	Mean Square	Expected Mean Square
Between Row (Animal) Means	$r-1$	$S(X_{i.}-X_{..})^2$	$S(X_{i.}-X_{..})^2/(r-1)$	$\sigma^2 + c\sigma_r^2$
Between Column (Day) Means	$c-1$	$S(X_{.j}-X_{..})^2$	$S(X_{.j}-X_{..})^2/(c-1)$	$\sigma^2 + r\sigma_c^2$
Residual	$(r-1)(c-1)$	$S(X_{ij}-X_{i.}-X_{.j}+X_{..})^2$	$S(X_{ij}-X_{i.}-X_{.j}+X_{..})^2/(r-1)(c-1)$	σ^2
Total	$rc-1$	$S(X_{ij}-X_{..})^2$	$S(X_{ij}-X_{..})^2/(rc-1)$	$\sigma^2 + \frac{c(r-1)}{rc-1}\sigma_r^2 + \frac{r(c-1)}{rc-1}\sigma_c^2$

In summary, when *Assumptions A, B, and C,* or their analogs in more complex cases, are satisfied, the customary analysis-of-variance procedures yield unbiased estimates of the respective variance components: Details of the procedures appropriate to situations differing in various ways from the situation considered here will be found in the papers by S. Lee Crump, and by H. E. Daniels cited in the list of references; and in the additional references that they cite.

Whereas *Assumptions A, B, and C,* or their analogs in more complex cases, are necessary[10] and sufficient for the validity of analysis-of-

[10] See footnote 2.

variance procedures for *unbiased estimation* of *components of variance,* it is not possible to conduct exact tests of significance with respect to these components of variance, nor to derive exact confidence limits for them or their ratios, unless the joint distribution of the several *deviations* in relations (11) is specified. Therefore, we shall make

Assumption D: The deviations $(m_{i.} - m_{..})$, $(m_{.j} - m_{..})$, and z_{ij}, $(i = 1, 2, \ldots, r; j = 1, 2, \ldots, c)$, are all normally distributed.

When *Assumption D* is satisfied, *Assumptions A* and *B* are partially redundant, and serve principally to define the "compositions" of the random variables x_{ij}, $(i = 1, 2, \ldots r, j = 1, 2, \ldots, c)$. Furthermore, zero covariances, as postulated in *Assumption C,* taken in conjunction with normality, postulated in *Assumption D,* imply mutual independence of the deviations $(m_{i.} - m_{..})$, $(m_{.j} - m_{..})$, and z_{ij}; and thence of the x_{ij} with respect to each other. So, once again, independence gets in by the back door.

When *Assumptions A–D* are all satisfied, then all of the standard analysis-of-variance procedures for estimating, and testing to determine whether to infer the existence of, *components of variance* are strictly valid. These are based on that fact that these assumptions are sufficient to insure that

(a) The quotient (Between-row-means *sum* of squares)$/(\sigma^2 + c_r{}^2)$ will be distributed as χ^2 for $(r-1)$ degrees of freedom,

(b) The quotient (Between-column-means *sum of squares*)$/(\sigma^2 + r\sigma_c{}^2)$ will be distributed as χ^2 for $(c-1)$ degrees of freedom,

(c) The quotient (Residual *sum of squares*)$/\sigma^2$ will be distributed as χ^2 for $(r-1)(c-1)$ degrees of freedom,

(d) The "quotients" referred to in (a), (b), and (c) will be independent in the probability sense, so that

(e) The quantity

$$\left[\frac{\text{(Between-row-means } \textit{mean square}\text{)}}{\sigma^2 + c\sigma_r{}^2}\right] \Big/ \left[\frac{\text{Residual } \textit{mean square}}{\sigma^2}\right]$$

will be distributed in Snedecor's F-distribution for $n_1 = (r-1)$ and $n_2 = (r-1)(c-1)$ degrees of freedom, and

(f) The quantity

$$\left[\frac{\text{Between-column-means } \textit{mean square}}{\sigma^2 + r\sigma_c{}^2}\right] \Big/ \left[\frac{\text{Residual } \textit{mean square}}{\sigma^2}\right]$$

will be distributed according F for $n_1 = (c-1)$ and $n_2 = (r-1)(c-1)$ degrees of freedom.

Thus (c), which obtains also in the case of Model I when *Assumptions 1–4* are satisfied, is the basis of exact tests of hypotheses regarding the value of σ^2, and of the derivation of exact confidence limits for the value of σ^2. Similarly (e) is the basis of exact tests of hypotheses regarding the value of σ_r^2/σ^2, e.g., that $\sigma_r^2 = 0$, and of the derivation of exact confidence limits for σ_r^2/σ^2. An analogous statement holds for (f) in relation to σ_c^2/σ^2.[11]

Unfortunately, aside from testing the hypothesis that $\sigma_r^2 = 0$ or that $\sigma_c^2 = 0$, it is not possible to conduct exact tests of hypotheses regarding the absolute values of σ_r^2 and σ_c^2, nor is it possible to derive exact confidence limits for their absolute values.

5. *Which Model—Model I or Model II?* In practical work a question that often arises is: which model is appropriate in the present instance—Model I or Model II? Basically, of course, the answer is clear as soon as a decision is reached on whether the parameters of interest specify *fixed relations*, or *components of random variation.* The answer depends in part, however, upon how the observations were obtained; on the extent to which the experimental procedure employed sampled the respective variables at random. This generally provides the clue. For instance, when an experimenter selects two or more treatments, or two or more varieties, for testing, he rarely, if ever, draws them at random from a population of possible treatments or varieties; he selects those that he believes are most promising. Accordingly Model I is generally appropriate where treatment, or variety comparisons are involved. On the other hand, when an experimenter selects a sample of animals from a herd or a species, for a study of the effects of various treatments, he can insure that they are a random sample from the herd, by introducing randomization into the sampling procedure, for example, by using a table of random numbers. But he may consider such a sample to be a random sample from the species, only by making the assumption that the herd itself is a random sample from the species. In such a case, if several herds (from the same species) are involved, Model II would clearly be appropriate with respect to the variation among the animals from each of the respective herds, and might be appropriate with respect to the variation of the herds from one another.

[11] For detailed considerations of various aspects of planning and interpreting experiments for comparing standard deviations and components of variance, the reader is referred to the report by A. H. J. Baines and to Chapter 8 of the forthcoming book by the Statistical Research Group, Columbia University, which are cited in the list of references at the end of this paper.

The most difficult decisions are usually associated with *places* and *times:* Are the *fields* on which the tests were conducted a random sample of the county, or of the state, etc.? Are the *years* in which the tests were conducted a random sample of years?

When a particular experiment is being planned, or when the results are in and are being interpreted, the following parallel sets of questions serve to focus attention on the pertinent issues, and have been found helpful in answering the basic question of random versus fixed effects:

(1) Are the conclusions to be confined to the things actually studied (the animals, or the plots); to the immediate sources of these things (the herds, or the fields); or expanded to apply to more general populations (the species, or the farmland of the state)?

(2) In complete repetitions of the experiment would the same things be studied again (the same animals, or the same plots); would new samples be drawn from the identical sources (new samples of animals from the same herds, or new experimental arrangements on the same fields); or would new samples be drawn from the more general populations (new samples of animals from new herds, or new experimental arrangements on new fields)?

It is hoped that these queries will not only aid in reducing the reader's "headaches," but will lead him to the correct decisions.

Finally, it needs to be said—as the reader will no doubt discover for himself, when he considers some specific sets of data or some proposed experiments in the light of the above queries—that real-life investigations rarely fall entirely within the domain of Model I, or entirely within the domain of Model II, unless they are planned and conducted so as to achieve one or the other of these objectives, and then they may not be realistic. In consequence, some of the mean squares of the analysis-of-variance tables may be unbiased estimators of linear combinations of variance components; and others, of linear combinations of variance components *and* "mean squares" of *fixed deviations.* H. E. Daniels, in the paper cited in the list of references, has proposed a method of interpreting analysis-of-variance tables of this sort. His method consists essentially of looking at such an analysis-of-variance table through Model-II spectacles, and interpreting the "mean squares" of *fixed deviations* as variance components also. While this approach may be fruitful in situations of the type to which he has applied his method, it cannot be regarded as a general solution since, the objectives of problems of Class I and problems of Class II are in general quite distinct. More general methods need to be devised for interpreting "mixed" analysis-of-variance tables, particularly in regard to tests of significance for individual factors.

REFERENCES

Baines, A. H. J. *On Economical Design of Statistical Experiments.* (British) Ministry of Supply, Advisory Service on Statistical Method and Quality Control, Technical Report, Series R, No. QC/R/15, July 15, 1944.

Cramér, H. *Mathematical Methods of Statistics.* Princeton University Press, Princeton, New Jersey. 1946.

Crump, S. Lee. "The Estimation of Variance Components in Analysis of Variance," *Biometrics Bulletin,* vol. 2 (1946), pp. 7–11.

Daniels, H. E. "The Estimation of Components of Variance," *Supplement to the Journal of the Royal Statistical Society,* vol. 6 (1939), pp. 186–197.

David F. N., and Neyman, J. "Extension of the Markoff Theorem on Least Squares," *Statistical Research Memoirs,* vol. 2 (1938), pp. 105–116.

Fisher, R. A. *Statistical Methods for Research Workers,* 1st and later editions. Oliver & Boyd, Ltd., London and Edinburgh, 1925–1944.

Goulden, C. H. *Methods of Statistical Analysis.* John Wiley and Sons, New York. 1939.

Irwin, J. O. "Mathematical Theorems Involved in the Analysis of Variance," *Journal of the Royal Statistical Society,* vol. 94 (1931), pp. 285–300.

Irwin, J. O. "On the Independence of the Constituent Items in the Analysis of Variance," *Supplement to the Journal of the Royal Statistical Society,* vol. 1 (1934), pp. 236–251.

Kendall, M. G. *The Advanced Theory of Statistics,* Volume II. Charles Griffin & Co., Ltd., London. 1946.

Kolodziejczyk, S. "On an Important Class of Statistical Hypotheses," *Biometrika,* vol. 27 (1935), pp. 161–190.

Lehmer, Emma. "Inverse Tables of Probabilities of Errors of the Second Kind," *Annals of Mathematical Statistics,* vol. 15 (1944), pp. 388–398.

Snedecor, G. W. *Calculation and Interpretation of Analysis of Variance and Covariance.* The Collegiate Press, Inc., Ames, Iowa. 1934.

Snedecor, G. W. *Statistical Methods: Applied to Experiments in Agriculture and Biology,* 4th Edition. The Collegiate Press, Inc., Ames, Iowa. 1946.

Statistical Research Group, Columbia University. *Selected Techniques of Statistical Analysis: For Scientific and Industrial Research and Production and Management Engineering.* McGraw-Hill Book Company, Inc., New York. (In press.)

Tang, P. C. "The Power Function of the Analysis of Variance Tests with Tables and Illustrations of Their Use," *Statistical Research Memoirs,* vol. 2 (1938), pp. 126–149.

Tippett, L. H. C. *The Methods of Statistics,* 2nd Edition. Williams and Norgate, Ltd., London. 1937.

Wilks, S. S. *Mathematical Statistics.* Princeton University Press, Princeton, New Jersey. 1943.

SOME CONSEQUENCES WHEN THE ASSUMPTIONS FOR THE ANALYSIS OF VARIANCE ARE NOT SATISFIED

W. G. Cochran

Institute of Statistics, North Carolina State College

1. *Purposes of the Analysis of Variance.* The main purposes are:

(i) To estimate certain treatment differences that are of interest. In this statement both the words "treatment" and "difference" are used in rather a loose sense: e.g., a treatment difference might be the difference between the mean yields of two varieties in a plant-breeding trial, or the relative toxicity of an unknown to a standard poison in a dosage-mortality experiment. We want such estimates to be *efficient.* That is, speaking roughly, we want the difference between the estimate and the true value to have as small a variance as can be attained from the data that are being analyzed.

(ii) To obtain some idea of the accuracy of our estimates, e.g., by attaching to them estimated standard errors, fiducial or confidence limits, etc. Such standard errors, etc., should be reasonably free from bias. The usual property of the analysis of variance, when all assumptions are fulfilled, is that estimated variances are unbiased.

(iii) To perform tests of significance. The most common are the F-test of the null hypothesis that a group of means all have the same true value, and the t-test of the null hypothesis that a treatment difference is zero or has some known value. We should like such tests to be *valid,* in the sense that if the table shows a significance probability of, say, 0.023, the chance of getting the observed result or a more discordant one on the null hypothesis should really be 0.023 or something near it. Further, such tests should be *sensitive* or *powerful,* meaning that they should detect the presence of real treatment differences as often as possible.

The object of this paper is to describe what happens to these desirable properties of the analysis of variance when the assumptions required for the technique do not hold. Obviously, any practical value of the paper will be increased if advice can also be given on how to detect failure of the assumptions and how to avoid the more serious consequences.

2. *Assumptions Required for the Analysis of Variance.* In setting up an analysis of variance, we generally recognize three types of effect:

(a) treatment effects—the effects of procedures deliberately introduced by the experimenter

(b) environmental effects (the term is not ideal)—these are certain features of the environment which the analysis enables us to measure. Common examples are the effects of replications in a randomized blocks experiment, or of rows and columns in a Latin square

(c) experimental errors—this term includes all elements of variation that are not taken account of in (a) or (b).

The assumptions required in the analysis of variance for the properties listed as desirable in section 1 are as follows:

(1) The treatment effects and the environmental effects must be additive. For instance, in a randomized blocks trial the observation y_{ij} on the i^{th} treatment in the j^{th} replication is specified as

$$y_{ij} = \mu + \tau_i + \rho_j + e_{ij}$$

where μ is the general mean, τ_i is the effect of the i^{th} treatment, ρ_j is the effect of the j^{th} replication and e_{ij} is the experimental error of that observation. We may assume, without loss of generality, that the e's all have zero means.

(2) The experimental errors must all be independent. That is, the probability that the error of any observation has a particular value must not depend on the values of the errors for other observations.

(3) The experimental errors must have a common variance.[1]

(4) The experimental errors should be normally distributed.

We propose to consider each assumption and to discuss the consequences when the assumption is not satisfied. The discussion will be in rather general terms, for much more research would be needed in order to make precise statements. Moreover, in practice several assumptions may fail to hold simultaneously. For example, in non-normal distributions there is usually a correlation between the variance of an observation and its mean, so that failure of condition (4) is likely to be accompanied by failure of (3) also.

3. *Previous Work on the Effects of Non-normality.* Most of the published work on the effects of failures in the assumptions has been

[1] This statement, though it applies to the simplest analyses, is an oversimplification. More generally, the analysis of variance should be divisible into parts within each of which the errors have common variance. For instance, in the split-plot design, we specify one error variance for whole-plot comparisons and a different one for subplot comparisons.

concerned with this item. Writing in 1938, Hey (8) gives a bibliography of 36 papers, most of which deal with non-normality, while several theoretical investigations were outside the scope of his bibliography. Although space does not permit a detailed survey of this literature, some comments on the nature of the work are relevant.

The work is almost entirely confined to a single aspect, namely the effect on what we have called the validity of tests of significance. Further, insofar as the t-test is discussed, this is either the test of a single mean or of the difference between the means of two groups. As will be seen later, it is important to bear this restriction in mind when evaluating the scope of the results.

Some writers, e.g., Bartlett (1), investigated by mathematical methods the theoretical frequency distribution of F or t, assuming the null hypothesis true, when sampling from an infinite population that was non-normal. As a rule, it is extremely difficult to obtain the distributions in such cases. Others, e.g., E. S. Pearson (9), drew mechanically 500 or 1000 numerical samples from an infinite non-normal population, calculated the value of F or t for each sample, and thus obtained empirically some idea of their frequency distributions. Where this method was used, the number of samples was seldom large enough to allow more than a chi-square goodness of fit test of the difference between the observed and the standard distributions. A very large number of samples is needed to determine the 5 percent point, and more so the 1 percent point, accurately. A third method, of which Hey's paper contains several examples, is to take actual data from experiments and generate the F or t distribution by means of randomization similar to that which would be practiced in an experiment. The data are chosen, of course, because they represent some type of departure from normality.

The consensus from these investigations is that no serious error is introduced by non-normality in the significance levels of the F-test or of the two-tailed t-test. While it is difficult to generalize about the range of populations that were investigated, this appears to cover most cases encountered in practice. If a guess may be made about the limits of error, the true probability corresponding to the tabular 5 percent significance level may lie between 4 and 7 percent. For the 1 percent level, the limits might be taken as $\frac{1}{2}$ percent and 2 percent. As a rule, the tabular probability is an underestimate: that is, by using the ordinary F and t tables we tend to err in the direction of announcing too many significant results.

The one-tailed *t*-test is more vulnerable. With a markedly skew distribution of errors, where one tail is much longer than the other, the usual practice of calculating the significance probability as one-half the value read from the tables may give quite a serious over- or under-estimate.

It was pointed out that work on the validity of the *t*-test covered only the cases of a single mean or of the comparison of the means of two groups. The results would be applicable to a randomized blocks experiment if we adopted the practice of calculating a separate error for each pair of treatments to be tested, using only the data from that pair of treatments. In practice, however, it is usual to employ a pooled error for all *t*-tests in an analysis, since this procedure not only saves labor but provides more degrees of freedom for the estimation of error. It will be shown in section 6 that this use of a pooled error when non-normality is present may lead to large errors in the significance probabilities of individual *t*-tests. The same remark applies to the Latin square and more complex arrangements, where in general it is impossible to isolate a separate error appropriate to a given pair of treatments, so that pooling of errors is unavoidable.

4. *Further Effects of Non-Normality.* In addition to its effects on the validity of tests of significance, non-normality is likely to be accompanied by a loss of efficiency in the estimation of treatment effects and a corresponding loss of power in the *F*- and *t*-tests. This loss of efficiency has been calculated by theoretical methods for a number of types of non-normal distribution. While these investigations dealt with the estimation of a single mean, and thus would be strictly applicable only to a paired experiment analyzed by the method of differences, the results are probably indicative of those that would be found for more complex analyses. In an attempt to use these results for our present purpose, the missing link is that we do not know which of the theoretical non-normal distributions that have been studied are typical of the error distributions that turn up in practice. This gap makes speculation hazardous, because the efficiency of analysis of variance methods has been found to vary from 100 percent to zero. While I would not wish to express any opinion very forcibly, my impression is that in practice the loss of efficiency is not often great. For instance, in an examination of the Pearson curves, Fisher (4) has proved that for curves that exhibit only a moderate departure from normality, the efficiency remains reasonably high. Further, an analysis of the logs of the observations instead of the observations themselves has fre-

quently been found successful in converting data to a scale where errors are approximately normally distributed. In this connection, Finney, (3) has shown that if log x is exactly normally distributed, the arithmetic mean of x has an efficiency greater than 93 percent so long as the coefficient of variation of x is less than 100 percent. In most lines of work a standard error as high as 100 percent per observation is rare, though not impossible.

The effect of non-normality on estimated standard errors is analogous to the effect on the t-test. If a standard error is calculated specifically for each pair of treatments whose means are to be compared, the error variance is unbiased. Bias may arise, however, by the application of a pooled error to a particular pair of treatments.

We now consider how to detect non-normality. It might perhaps be suggested that the standard tests for departure from normality, Fisher (5), should be applied to the errors in an analysis. This suggestion is not fruitful, however, because for experiments of the size usually conducted, the tests would detect only very violent skewness or kurtosis. Moreover, as is perhaps more important, it is not enough to detect non-normality: in order to develop an improved analysis, one must have some idea of the actual form of the distribution of errors, and for this purpose a single experiment is rarely adequate.

Examination of the distribution of errors may be helpful where an extensive uniformity trial has been carried out, or where a whole series of experiments on similar material is available. Theoretically, the best procedure would be to try to find the form of the frequency distribution of errors, using, of course, any *a priori* knowledge of the nature of the data. An improved method of estimation could then be developed by maximum likelihood. This, however, would be likely to lead to involved computations. For that reason, the usual technique in practice is to seek, from *a priori* knowledge or by trial and error, a transformation that will put the data on a scale where the errors are approximately normal. The hope is that in the transformed scale the usual analysis will be reasonably efficient. Further, we would be prepared to accept some loss in efficiency for the convenience of using a familiar method. Since a detailed account of transformations will be given by Dr. Bartlett in the following paper, this point will not be elaborated.

The above remarks are intended to apply to the handling of a rather extensive body of data. With a single experiment, standing by itself, experience has indicated two features that should be watched for:

(i) evidence of charges in the variance from one part of the experiment to another. This case will be discussed in section 6.

(ii) evidence of gross errors.

5. *Effects of Gross Errors.* The effects of gross errors, if undetected, are obvious. The means of the treatments that are affected will be poorly estimated, while if a pooled error is used the standard errors of other treatment means will be over-estimated. An extreme example is illustrated by the data in Table I, which come from a randomized blocks experiment with four replicates.

TABLE I

WHEAT: RATIO OF DRY TO WET GRAIN

Block	Nitrogen applied			
	None	Early	Middle	Late
1	.718	.732	.734	.792
2	.725	.781	.725	.716
3	.704	1.035	.763	.758
4	.726	.765	.738	.781

As is likely to happen when the experimenter does not scrutinize his own data, the gross error was at first unnoticed when the computor carried out the analysis of variance, though the value is clearly impossible from the nature of the measurements. This fact justifies rejection of the value and substitution of another by the method of missing plots, Yates (11).

Where no explanation can be found for an anomalous observation, the case for rejection is more doubtful. Habitual rejection of outlying values leads to a marked underestimation of errors. An approximate test of significance of the contribution of the suspected observation to the error helps to guard against this bias. First calculate the error sum of squares from the actual observations. Then calculate the error when the suspected value is replaced by the missing-plot estimate: this will have one less degree of freedom and is designated the "Remainder" in the data below. The difference represents the sum of squares due to the suspect. For the data above, the results are

	d.f.	S.S.	M.S.
Actual error	9	.04729	.00525
Suspect	1	.04205	.04205
Remainder	8	.00524	.000655

Alternatively, the contribution due to the suspected observation may be calculated directly and the remainder found by subtraction. If there are t treatments and r replicates, the sum of squares is $(t-1)(r-1)d^2/tr$, where d is the difference between the suspected observation and the value given by the missing-plot formula. In the present case t and r are 3 and the missing-plot value is 0.7616, so that the contribution is $9(0.2734)^2/16$, or 0.04205.[2]

The F ratio for the test of the suspect against the remainder is 64.2, giving a t value of 8.01, with 8 degrees of freedom. Now, assuming that the suspect had been examined simply because it appeared anomalous, with no explanation for the anomaly, account must be taken of this fact in the test of significance. What is wanted is a test appropriate to the *largest* contribution of any observation. Such a test has not as yet been developed. The following is suggested as a rough approximation. Calculate the significance probability, p, by the ordinary t table. Then use as the correct significance probability np, where n is the number of degrees of freedom in the actual error.[3] In the present case, with $t = 8.01$, p is much less than 1 in a million, and consequently np is less than 1 in 100,000. In general, it would be wise to insist on a rather low significance probability (e.g., 1 in 100) before rejecting the suspect, though a careful answer on this point requires knowledge of the particular types of error to which the experimentation is subject.

6. *Effects of Heterogeneity of Errors.* If ordinary analysis of variance methods are used when the true error variance differs from one observation to another, there will as a rule be a loss of efficiency in the estimates of treatment effects. Similarly, there will be a loss of sensitivity in tests of significance. If the changes in the error variance are large, these losses may be substantial. The validity of the F-test for all treatments is probably the least affected. Since, however, some treatment comparisons may have much smaller errors than others, t-tests from a pooled error may give a serious distortion of the significance levels. In the same way the standard errors of particular treatment comparisons, if derived from a pooled error, may be far from the true values.

[2] This formula applies only to randomized blocks. Corresponding formulas can be found for other types of arrangements. For instance, the formula for a $p \times p$ Latin square is $(p-1)(p-2)d^2/p^2$.

[3] The approximation is intended only to distinguish quickly whether the probability is low or high and must not be regarded as accurate. For a discussion of this type of test in a somewhat simpler case, see E. S. Pearson and C. Chandra Sekar, *Biometrika,* Vol. 28 (1936), pp. 308–320.

There is no theoretical difficulty in extending the analysis of variance so as to take account of variations in error variances. The usual analysis is replaced by a weighted analysis in which each observation is weighted in proportion to the inverse of its error variance. The extension postulates, however, a knowledge of the relative variances of any two observations and this knowledge is seldom available in practice. Nevertheless, the more exact theory can sometimes be used with profit in cases where we have good estimates of these relative variances. Suppose for instance, the situation were such that the observations could be divided into three parts, the error variances being constant within each part. If unbiased estimates of the variances within each part could be obtained and if these were each based on, say, at least 15 degrees of freedom, we could recover most of the loss in efficiency by weighting inversely as the observed variances. This device is therefore worth keeping in mind, though in complex analyses the weighted solution involves heavy computation.

TABLE II

MANGOLDS, PLANT NUMBERS PER PLOT

Block	Control		Chalk			Lime			Total
	0	0	1	2	3	1	2	3	
I	140	49	98	135	117	81	147	130	897
II	142	37	132	151	137	129	131	112	971
III	36	114	130	143	137	135	103	130	928
IV	129	125	153	146	143	104	147	121	1068
Total	447	325	513	575	534	449	528	493	3864
Range	106	88	55	16	26	54	44	18	

Heterogeneity of errors may arise in several ways. It may be produced by mishaps or damage to some part of the experiment. It may be present in one or two replications through the use of less homogeneous material or of less carefully controlled conditions. The nature of the treatments may be such that some give more variable responses than others. An example of this type is given by the data in Table II.

The experiment investigated the effects of three levels of chalk dressing and three of lime dressing on plant numbers of mangolds. There were four randomized blocks of eight plots each, the control plots being replicated twice within each block.[4]

Since the soil was acid, high variability might be anticipated for the

[4] The same data were discussed (in much less detail) in a previous paper, Cochran (2).

control plots as a result of partial failures on some plots. The effect is evident on eye inspection of the data. To a smaller extent the same effect is indicated on the plots receiving the single dressing of chalk or lime. If the variance may be regarded as constant within each treatment, there will be no loss of efficiency in the treatment means in this case, contrary to the usual effect of heterogeneity. Any *t* tests will be affected and standard errors may be biased. In amending the analysis so as to avoid such disturbances, the first step is to attempt to subdivide the error into homogeneous components. The simple analysis of variance is shown below.

TABLE III

Analysis of Variance for Mangolds Data

	d.f.	S.S.	M.S.
Blocks	3	2,079	....
Treatments	6	8,516	....
Error	22	18,939	860.9
Total	31	29,534	....

For subdivision of the error we need the following auxiliary data.

Block	Diff. between Controls	Total – 4 (Controls)	(C1–L1)	(C2 + L2 + C3 + L3) – 2(C1 + L1)
1	91	141	17	171
2	105	255	3	9
3	78	328	–5	–17
4	4	52	49	43
Total	....	776	64	206
Divisor for S.S.	2	24	2	12

The first two columns are used to separate the contribution of the controls to the error. This has 7 d.f. of which 4 represent differences between the two controls in each block. The sum of squares of the first column is divided by 2 as indicated. There remain 3 d.f. which come from a comparison within each block of the total yield of the controls with the total yield of the dressings. Since there are 6 dressed plots to 2 controls per block we take

$$(\text{Dressing total}) - 3(\text{Control total}) = (\text{Total}) - 4(\text{Control total})$$

Thus $141 = 897 - 4(140 + 49)$.

By the usual rule the divisor for the sum of squares of deviations is 24.

Two more columns are used to separate the contribution of the single dressings. There are 6 d.f. of which 3 compare chalk with line at this level while the remaining 3 compare the single level with the higher levels. The resulting partition of the error sum of squares is shown below.

TABLE IV

PARTITION OF ERROR SUM OF SQUARES

	d.f.	S.S.	M.S.
Total	22	18,939	861
Between controls	4	12,703	3,176
Controls v. Dressings	3	1,860	620
Chalk 1 v. Lime 1	3	850	283
Single v. Higher Dressings	3	1,738	579
Double and Triple Dressings	9	1,788	199

As an illustration of the disturbance to *t*-tests and to estimated standard errors, we may note that the pooled mean square, 861, is over four times as large as the 9 d.f. error, 199, obtained from the double and triple dressings. Consequently, the significance levels of *t* and standard errors would be inflated by a factor of two if the pooled error were applied to comparisons within the higher dressings.

In a more realistic approach we might postulate three error variances, σ_c^2 for controls, σ_1^2 for single dressings and σ_h^2 for higher dressings. For these we have unbiased estimates of 3,176, 283 and 199 respectively from Table IV. The mean square for Controls v. Dressings (620) would be an unbiased estimate of $(9\sigma_c^2+\sigma_1^2+2\sigma_h^2)/12$, while that for Single v. Higher Dressings (579) would estimate $(2\sigma_1^2+\sigma_h^2)/3$.

What one does in handling comparisons that involve different levels depends on the amount of refinement that is desired and the amount of work that seems justifiable. The simplest process is to calculate a separate *t*-test or standard error for any comparison by obtaining the comparison separately within each block. Such errors, being based on 3 d.f., would be rather poorly determined. A more complex but more efficient approach is to estimate the three variances from the five mean squares given above. Since the error variance of any comparison will be some linear function of these three variances, it can then be estimated.

To summarize, heterogeneity of errors may affect certain treatments or certain parts of the data to an unpredictable extent. Sometimes, as in the previous example, such heterogeneity would be expected in advance from the nature of the experiment. In such cases the data may be inspected carefully to decide whether the actual amount of variation in the error variance seems enough to justify special methods. In fact, such inspection is worthwhile as a routine procedure and is, of course, the only method for detecting heterogeneity when it has not been anticipated. The principal weapons for dealing with this irregular type of heterogeneity are subdivision of the error variance or omission of parts of the experiment. Unfortunately, in complex analyses the computations may be laborious. For the Latin square, Yates (12) has given methods for omitting a single treatment, row or column, while Yates and Hale (14) have extended the process to a pair of treatments, rows or columns.

In addition, there is a common type of heterogeneity that is more regular. In this type, which usually arises from non-normality in the distribution of errors, the variance of an observation is some simple function of its mean value, irrespective of the treatment or block concerned. For instance, in counts whose error distribution is related to the Poisson, the variance of an observation may be proportional to its mean value. Such cases, which have been most successfully handled by means of transformations, are discussed in more detail in Dr. Bartlett's paper.

7. *Effects of Correlations Amongst the Errors.* These effects may be illustrated by a simple theoretical example. Suppose that the errors $e_1, e_2, \ldots, e_r$ of the r observations on a treatment in a simple group comparison have constant variance σ^2 and that every pair has a correlation coefficient ρ. The error of the treatment total, $(e_1 + e_2 \ldots, + e_r)$ will have a variance

$$r\sigma^2 + r(r-1)\rho\sigma^2$$

since there are $r(r-1)/2$ cross-product terms, each of which will contribute $2\rho\sigma^2$. Hence the *true* variance of the treatment mean is

$$\sigma^2\{1 + (r-1)\rho\}/r.$$

Now in practice we would estimate this variance by means of the sum of squares of deviations within the group, divided by $r(r-1)$. But

$$\text{Mean } \Sigma(e_i - \bar{e})^2 = \text{Mean } \Sigma\, e_i{}^2 - r\,\{\text{Mean } \bar{e}^2\}$$
$$= r\sigma^2 - \sigma^2\{1 + (r-1)\rho\} = (r-1)\sigma^2(1-\rho).$$

Hence the *estimated* variance of the treatment mean is $\sigma^2(1-\rho)/r$.

Consequently, if ρ is positive the treatment mean is less accurate than the mean of an independent series, but is estimated to be more accurate. If ρ is negative, these conditions are reversed. Substantial biases in standard errors might result, with similar impairment of t-tests. Moreover, in many types of data, particularly field experimentation, the observations *are* mutually correlated, though in a more intricate pattern.

Whatever the nature of the correlation system, this difficulty is largely taken care of by proper randomization. While mathematical details will not be given, the effect of randomization is, roughly speaking, that we may treat the errors as if they were independent. The reader may refer to a paper by Yates (13), which presents the nature of this argument, and to papers by Bartlett (1), Fisher (6) and Hey (8), which illustrate how randomization generates a close approximation to the F and t distributions.

Occasionally it may be discovered that the data have been subject to some systematic pattern of environmental variation that the randomization has been unable to cope with. If the environmental pattern obviously masks the treatment effects, resort may be had to what might be called desperate remedies in order to salvage some information.

The data in Table V provide an instance. The experiment was a 2^4 factorial, testing the effects of lime (L), fish manure (F) and artificial fertilizers (A). Lime was applied in the first year only; the other dressings were either applied in the first year only (1) or at a half rate every year (2). Two randomized blocks were laid out, the crop being pyrethrum, which forms an ingredient in many common insecti-

TABLE V

Weights of Dry Heads per Plot
(Unit, 10 grams)

Block 1				Block 2			
LA1	LF2	F2	L1	A1	L1	A2	0
84	66	70	81	63	97	56	64
1	1	1	1	1	1	1	1
LF1	A2	A1	FA2	F1	LA2	LA1	LFA1
148	137	146	171	168	158	189	152
0	0	0	0	0	0	0	0
LFA2	F1	LFA1	LA2	LF1	L2	LF2	FA2
179	218	247	228	191	195	189	179
0	0	0	0	0	0	0	0
0	L2	0	FA1	FA1	LFA2	0	F2
124	166	177	153	133	145	141	130
0	0	0	0	0	0	0	0

cides. The data presented are for the fourth year of the experiment, which was conducted at the Woburn Experimental Farm, England.

The weights of dry heads are shown immediately underneath the treatment symbols. It is evident that the first row of plots is of poor fertility—treatments appearing in that row have only about half the yields that they give elsewhere. Further, there are indications that every row differs in fertility, the last row being second worst and the third row best. The fertility gradients are especially troublesome in that the four untreated controls all happen to lie in outside rows. The two replications give practically identical totals and remove none of this variation.

There is clearly little hope of obtaining information about the treatment effects unless weights are adjusted for differences in fertility from row to row. The adjustment may be made by covariance.

For simplicity, adjustments for the first row only will be shown: these remove the most serious environmental disturbance. As x variable we choose a variable that takes the value 1 for all plots in the first row and zero elsewhere. The x values are shown under the weights in Table V. The rest of the analysis follows the usual covariance technique, Snedecor (10).

TABLE VI

Sums of Squares and Products
(y = weights, x = dummy variates)

	d.f.	y^2	yx	x^2
Blocks	1	657	0.0	0.00
Treatments	13	33,323	−200.2	1.75
Error	17	46,486	−380.0	4.25
Total	31	80,466	−580.2	6.00

Note that there are only 14 distinct treatments, since L1 is the same as L2. The reduction in the error S.S. due to covariance is $(380.0)^2/4.25$, or 33,976. The error mean square is reduced from 2,734 to 782 by means of the covariance, i.e., to less than one-third of its original value. The regression coefficient is $-380.0/4.25$, or -89.4 units.

Treatment means are adjusted in the usual way. For L1, which was unlucky in having two plots in the first row, the unadjusted mean is 89. The mean x value is 1, whereas the mean x value for the whole experiment is 8/32, or $\frac{1}{4}$. Hence the adjustment increases the L1 mean by (3/4) (89.4), the adjusted value being 156. For L2, which had no plots in the first row, the x mean is 0, and the adjustment reduces the mean from 180 to 158. It may be observed that the unadjusted mean of L2 was double that of L1, while the two adjusted means agree closely, as is reasonable since the two treatments are in fact identical.

If it were desired to adjust separately for every row, a multiple covariance with four x variables could be computed. Each x would take the value 1 for all plots in the corresponding row and 0 elsewhere. It will be realized that the covariance technique, if misused, can lead to an underestimation of errors. It is, however, worth keeping in mind as an occasional weapon for difficult cases.

8. *Effects of Non-Additivity.* Suppose that in a randomized blocks experiment, with two treatments and two replicates, the treatment and block effects are multiplicative rather than additive. That is, in either replicate, treatment B exceeds treatment A by a fixed percentage, while for either treatment, replicate 2 exceeds replicate 1 by a fixed percentage. Consider treatment percentages of 20% and 100% and replicate percentages of 10% and 50%. These together provide four combinations. Taking the observation for treatment A in replicate 1 as 1.0, the other observations are shown in Table VII.

TABLE VII

Hypothetical Data for Four Cases Where Effects Are Multiplicative

Rep.	T 20% R 10%		T 20% R 50%		T 100% R 10%		T 100% R 50%	
	A	B	A	B	A	B	A	B
1 2	1.0 1.1	1.2 1.32	1.0 1.5	1.2 1.8	1.0 1.1	2.0 2.2	1.0 1.5	2.0 3.0
d	.02		.10		.10		.50	
σ_{na}	.01		.05		.05		.25	

Thus, in the first case, 1.32 for B in replicate 2 is 1.2 times 1.1. Since no experimental error has been added, the error variance in a correct analysis should be zero. If the usual analysis of variance is applied to each little table, the calculated error in each case will have 1 d.f. If d is the sum of two corners minus the other two corners, the error S.S. is $d^2/4$, so that the standard error σ_{na} is $d/2$ (taken as positive). The values of d and of σ_{na} are shown below each table.

Consequently, in the first experiment, say, the usual analysis would lead to the statement that the average increase to B is 0.21 units ± 0.01, instead of to the correct statement that the increase to B is 20%. The standard error, although due entirely to the failure of the additive relationship, does perform a useful purpose. It warns us that the actual increase to B over A will vary from replication to replication and measures how much it will vary, so far as the experiment is capable of supplying information on this point. An experimenter who fails to see the correct method of analysis and uses ordinary methods will get less precise information from the experiment for predictive purposes, but if he notes the standard error he will not be misled into thinking that his information is more precise than it really is.

When experimental errors are present, the variance σ_{na}^2 will be added to the usual error variance σ_e^2. The ratio $\sigma_{na}^2/(\sigma_{na}^2+\sigma_e^2)$ may appropriately be taken as a measure of the loss (fractional) of information due to non-additivity. In the four experiments, from left to right, the values of σ_{na} are respectively 0.9, 3.6, 3.2, and 13.3 percent of the mean yields of the experiments. In the first case, where treatment and replicate effects are small, the loss of information due to non-additivity will be trivial unless σ_e is very small. For example, with $\sigma_e = 5$ percent, the fractional loss is 0.81/25.81 or about 3 percent. In the two middle examples, where either the treatment or the replicate effect is substantial, the losses are beginning to be substantial. With $\sigma_e = 5$ percent in the second case, the loss would be about 30 percent. Finally, when both effects are large the loss is great.

Little study has been made in the literature of the general effects of non-additivity or of the extent to which this problem is present in the data that are usually handled by analysis of variance.[5] I believe, however, that the results from these examples are suggestive of the consequences in other cases. The principal effect is a loss of information. Unless experimental errors are low or there is a very serious departure from additivity, this loss should be negligible when treatment and replication effects do not exceed 20 percent, since within that range the additive relationship is likely to be a good approximation to most types that may arise.

Since the deviations from additivity are, as it were, amalgamated with the true error variance, the pooled error variance as calculated from the analysis of variance will take account of these deviations and should be relatively unbiased. This pooled variance may not, however, be applicable to comparisons between individual pairs of treatments. The examples above are too small to illustrate this point. But, clearly, with three treatments A, B, and C, the comparison (A–B) might be much less affected by non-additivity than the comparison (A–C).

[5] A relevant discussion of this problem for regressions in general, with some interesting results, has been given recently by Jones (7).

Thus non-additivity tends to produce heterogeneity of the error variance.[6]

If treatment or block effects, or both, are large, it will be worth examining whether treatment differences appear to be independent of the block means, or vice versa. There are, of course, limitations to what can be discovered from a single experiment. If relations seem non-additive, the next step is to seek a scale on which effects are additive. Again reference should be made to the paper following on transformations.

9. *Summary and Concluding Remarks.* The analysis of variance depends on the assumptions that the treatment and environmental effects are additive and that the experimental errors are independent in the probability sense, have equal variance and are normally distributed. Failure of any assumption will impair to some extent the standard properties on which the widespread utility of the technique depends. Since an experimenter could rarely, if ever, convince himself that all the assumptions were exactly satisfied in his data, the technique must be regarded as approximative rather than exact. From general knowledge of the nature of the data and from a careful scrutiny of the data before analysis, it is believed that cases where the standard analysis will give misleading results or produce a serious loss of information can be detected in advance.

In general, the factors that are liable to cause the most severe disturbances are extreme skewness, the presence of gross errors, anomalous behavior of certain treatments or parts of the experiment, marked departures from the additive relationship, and changes in the error variance, either related to the mean or to certain treatments or parts of the experiment. The principal methods for an improved analysis are the omission of certain observations, treatments, or replicates, subdivision of the error variance, and transformation to another scale before analysis. In some cases, as illustrated by the numerical examples, the more exact methods require considerable experience in the manipulation of the analysis of variance. Having diagnosed the trouble, the experimenter may frequently find it advisable to obtain the help of the mathematical statistician.

[6] It is an over-simplification to pretend, as in the discussion above, that the deviations from addivity act entirely like an additional component of random error. Discussion of the effects introduced by the systematic nature of the deviations would, however, unduly lengthen this paper.

REFERENCES

1. Bartlett, M. S. "The Effect of Non-Normality on the *t* Distribution," *Proceedings of the Cambridge Philosophical Society* (1935), 31, 223–231.
2. Cochran, W. G. "Some Difficulties in the Statistical Analysis of Replicated Experiments," *Empire Journal of Experimental Agriculture* (1938), 6, 157–175.
3. Finney, D. J. "On the Distribution of a Variate Whose Logarithm is Normally Distributed," *Journal of The Royal Statistical Society, Suppl.* (1941), 7, 155–161.
4. Fisher, R. A. "On the Mathematical Foundations of Theoretical Statistics," *Philosophical Transactions of the Royal Society of London*, A, 222 (1922), 309–368.
5. Fisher, R. A. *Statistical Methods for Research Workers.* Oliver and Boyd, Edinburgh, § 14.
6. Fisher, R. A. *The Design of Experiments.* Oliver and Boyd, Edinburgh, § 21.
7. Jones, H. L. "Linear Regression Functions with Neglected Variablies," *Journal of the American Statistical Association* (1946), 41, 356–369.
8. Hey, G. B. "A New Method of Experimental Sampling Illustrated on Certain Non-Normal Populations," *Biometrika* (1938), 30, 68–80.
9. Pearson, E. S. "The Analysis of Variance in Cases of Non-Normal Variation," *Biometrika* (1931), 23, 114.
10. Snedecor, G. W. *Statistical Methods.* Iowa State College Press, Ames, Ia. 4th ed. (1946). Chaps. 12 and 13.
11. Yates, F. "The Analysis of Replicated Experiments When the Field Results Are Incomplete," *Empire Journal of Experimental Agriculture* (1933), 1, 129–142.
12. Yates, F. "Incomplete Latin Squares," *Journal of Agricultural Science* (1936), 26, 301–315.
13. Yates, F. "The Formation of Latin Squares for Use in Field Experiments," *Empire Journal of Experimental Agriculture* (1933), 1, 235–244.
14. Yates, F., and Hale, R. W. "The Analysis of Latin Squares When Two or More Rows, Columns or Treatments Are Missing," *Journal of the Royal Statistical Society, Suppl.* (1939), 6, 67–79.

THE USE OF TRANSFORMATIONS

M. S. Bartlett

University of Cambridge, England, and University of North Carolina

1. *Theoretical Discussion.* The purpose of this note is to summarize the transformations which have been used on raw statistical data, with particular reference to analysis of variance. For any such analysis the usual purpose of the transformation is to change the scale of the measurements in order to make the analysis more valid. Thus the conditions required for assessing accuracy in the ordinary unweighted analysis of variance include the important one of a constant residual or error variance, and if the variance tends to change with the mean level of the measurements, the variance will only be stabilized by a suitable change of scale.

If the form of the change of variance with mean level is known, this determines the type of transformation to use. Suppose we write

$$\sigma_x{}^2 = f(m), \tag{1}$$

where $\sigma_x{}^2$ is the variance on the original scale of measurements x with the mean of x equal to m. Then for any function $g(x)$ we have approximately[1]

$$\sigma_g{}^2 = (dg/dm)^2 f(m), \tag{2}$$

so that if $\sigma_g{}^2$ is to be constant, C^2 say, we must have

$$g(m) = \int \frac{Cdm}{\sqrt{f(m)}}. \tag{3}$$

For example, if the standard deviation σ_x tends to be proportional to the mean level m, we have $f(m)$ proportional to m^2, and $g(m)$ proportional to log m; i.e., we should use the logarithmic scale. Appropriate scales of this kind for types of data often encountered in statistical analysis are discussed in sections 2, 3, and 4.

However, a constant variance is not the only condition we seek, and precautions are still necessary when using analysis of variance with the transformed variate. In the ideal case (cf. Reference 6 at the end of this paper),

(a) The variance of the transformed variate should be unaffected by changes in the mean level (this is taken to be the primary purpose of the transformations of sections 2, 3, and 4).

(b) The transformed variate should be normally distributed.

[1] For a more precise formulation, see Reference 15.

(c) The transformed scale should be one for which an arithmetic average is an efficient estimate of the true mean level for any particular group of measurements.

(d) The transformed scale should be one for which real effects are linear and additive.

Although these conditions are to some extent related [for example, (a) and (b) and (d) together imply (c)], we obviously cannot necessarily expect to arrange for conditions (b), (c), and (d) to be satisfied if our scale has already been fixed by condition (a).

Fortunately, the transformation of scale to meet condition (a) often has the effect of improving the closeness of the distribution to normality, a correlation of variability with mean level on the original scale often implying excessive skewness which tends to be eliminated after the transformation. But the validity of any assumption of normality should be watched, for while moderate departures from normality are known not to be serious, any large departures in the region of the more outlying observations are likely to affect the validity of significance tests (cf. Reference 6).

Condition (c) is required because the estimates which arise in analysis of variance are of the simple arithmetic average type, and we want to know that such estimates are efficient. The contention is sometimes made that the original scale is the more relevant one for taking sums and averages, and more understandable.

While this argument has some force and is a warning against making transformations without good reason, it loses strength when we remember that if the variability in the data varies with the mean level for different blocks or groups an unweighted average of the *observed* treatment responses is not necessarily the best estimate of the *true* treatment response, and the average on the transformed scale will often be the better estimate when re-converted to the original scale.

Lastly, it is a more effective and a simpler procedure to be working on a scale for which treatment or other effects are linear and additive; this implies in a layout of the randomized block type that real treatment × block interactions will not inflate the error term, and reduce the apparent significance of the treatments; and relatedly, that for treatments of the factorial type, interactions between the treatments due merely to scale will not necessitate narrower and less powerful conclusions about the treatment effects. Now it is not always possible to choose a scale to cover conditions (a) and yet be most reasonable for (d), though it may happen that a choice of scale for (a) improves the scale to some extent as far as (d) is concerned. In some cases where sufficient information about the variability is known from the nature of the data, we may decide to abandon the advantages of a simple analysis of variance under condition (a), and choose our scale with sole regard to (d), weighting our observations with appropriate weights depending on the known variability. Something like this happens, for example, when we make use of the probit transformation (section 5; cf. also section 7), which is chosen to provide a rational linear scale for percentage mortalities or other analogous percentages.

2. *The Square Root Transformation.* When statistical data consist of integers, i.e., whole numbers, such as number of bacterial colonies in a plate count, or number of plants in a given area, homogeneous conditions will often lead to variation in these numbers x following the Poisson distribution. Since for such a distribution the variance is exactly equal to the mean, we readily obtain from our general equation in section 1, that to stabilize the variance we must work on the square-root scale. We have seen that this is only an approximate result and the exact values for different values of the mean m seem worth quoting from my original paper (see Reference 1, Table I), as they may be useful in any comparisons of the observed variance of our data with this theoretical value. I recommended the use of $\sqrt{(x+\frac{1}{2})}$ in place of $\sqrt{x}$ when very small numbers were involved (e.g., means in the range 10 to 2, especially when zeros are occurring among the observed numbers), and the variance for this quantity is also shown.

In practice we often use an analysis of variance for data of the above integer type because we suspect heterogeneity of one kind or another to be present, especially if our data have been collected under field con-

TABLE I

Variance of Poisson Variate on Transformed Scale

Mean m (on original scale)	$\sqrt{x}$	$\sqrt{(x+\frac{1}{2})}$
0.0	0.000	0.000
0.5	0.310	0.102
1.0	0.402	0.160
2.0	0.390	0.214
3.0	0.340	0.232
4.0	0.306	0.240
6.0	0.276	0.245
9.0	0.263	0.247
12.0	0.259	0.248
15.0	0.256	0.248

ditions. We do not then need to assume Poisson variation, but will still transform to the square-root scale if the variation of $\sqrt{x}$ appears stable. The following set of data (quoted from Reference 1; see also Reference 13), representing weed-infestation counts in one of a series of experiments on weed control in cereals, is an example of stability of variance on the square-root scale, even when the level of variability is far higher than expected on the assumptions of a Poisson distribution for x.

TABLE II

PLAN SHOWING LAYOUT OF EXPERIMENT ON OATS, AND NUMBERS OF POPPIES (IN 3¾ SQ. FT. AREAS)

Block A	(1)* 438	(4) 17	(2) 538	(5) 18	(3) 77	(6) 115
B	(3) 61	(2) 422	(6) 57	(1)* 442	(5) 26	(4) 31
C	(5) 77	(3) 157	(4) 87	(6) 100	(2) 377	(1)* 319
D	(2) 315	(1)* 380	(5) 20	(3) 52	(4) 16	(6) 45

* Control.

As a rule, a scale chosen to stabilize variance will be one on which arithmetic averages will provide efficient estimates, though this requirement, which I have called condition (c), is not altogether independent of (d). For data for which "homogeneity" represents some well-defined assumption such as Poisson variation, it is also useful to be sure that if the data were in fact homogeneous, such estimates are not throwing away too much information, and this condition is satisfied for the Poisson variate transformed to the square-root scale (cf. Reference 1, p. 69, where it was noted that the minimum percentage efficiency in large samples was 88 percent for $\sqrt{x}$ and 96½ percent for $\sqrt{(x+\frac{1}{2})}$), even although the best estimate of the true mean of a perfectly homogeneous Poisson set of observations is actually the arithmetic mean on the original scale. In an interesting theoretical paper, Cochran (Reference 14) has discussed the appropriate analysis of variance for data of a Poisson (or binomial) type, for which real block or other group differences extra to the treatment effects may be present provided they are assumed additive on the transformed scale. He shows that the direct analysis of variance on the transformed scale is then really a first approximation to a more exact analysis, in which any loss of efficiency in estimation is reduced to zero. This method, however, becomes irrelevant if the data do not belong to the exact distributional type assumed, and consequently its use would seem rarely justifiable in practice for field data of the type here considered.

3. *Logarithmic Transformations.* The stability of variance on the square-root scale in the case of the series of weed-control experiments was rather unexpected; since, if considerable heterogeneity in numbers is present, the variance is often found still to be correlated with the mean level on a square-root scale, and may only be stabilized if transformation is made to the logarithmic scale. The natural explanation of a variance greater than the mean is that the mean level itself fluctuates, so that

$$\sigma_x^2 = m + \sigma_m^2. \tag{4}$$

For biological populations, increases in numbers are often proportional to the numbers already present, giving rise to variations in mean from place to place themselves proportional to the mean. This illustrates how σ_m^2 might be proportional to m^2, so that we might expect

$$\sigma_x^2 = m + \lambda^2 m^2. \tag{5}$$

For λ large, or m large, this variance law implies the logarithmic transmation.

In some problems it is possible that λ could be estimated well enough to justify a more exact transformation corresponding to a variance of the type represented by equation (5). This transformation would be λ^{-1} Sinh^{-1} $[\lambda\sqrt{x}]$, or equivalently λ^{-1} log $\{\sqrt{(1+\lambda^2 x)}+\lambda\sqrt{x}\}$ (cf. Reference 7). For example, it is known that under certain assumptions about the way m varies, the Poisson distribution becomes a "negative binomial" distribution, this distribution often fitting observational data which do not conform to the narrower Poisson type. For such data λ^{-1} Sinh^{-1} $[\lambda\sqrt{x}]$ transformation would be appropriate.[2] For small $\sqrt{x}$ it becomes equivalent to the $\sqrt{x}$ transformation, and for small numbers the transformation λ^{-1} Sinh^{-1} $[\lambda\sqrt{(x+\frac{1}{2})}]$ would seem somewhat better. For large $\lambda\sqrt{x}$ it becomes equivalent to the logarithmic transformation.

This transformation, however, has the disadvantage of requiring an approximate knowledge of λ, and the empirical transformation log $(1+x)$, which has been suggested in place of log x as a logarithmic transformation for integers to avoid the difficulty with zeros in the case

[2] Compare the Sin^{-1} $\sqrt{x}$ transformation for the ordinary binomial (section 4).

of log x, seems likely to prove good enough in many cases; (it shows an approximate linear relationship with Sinh^{-1} $[\lambda\sqrt{(x+\frac{1}{2})}]$ for values of λ which appear likely in practice). Beall (Reference 7) has, however, suggested that, in entomological field experiments where an estimate of λ is required, two plots for each treatment should be included in each randomized block; for such experimental designs the λ^{-1} Sinh^{-1} $[\lambda\sqrt{x}]$ scale would naturally be used.[3]

The Sinh^{-1} $[\lambda\sqrt{x}]$ scale is also appropriate for the somewhat more general variance law

$$\sigma_x^2 = \mu^2(m + \lambda^2 m^2). \qquad (6)$$

As an example[4] of data for which the empirical law (6) might have been fitted, the following "leatherjacket" counts (Table III) are cited. The figures refer to an experiment on the control of leatherjackets by the use of toxic emulsions (see Reference 2, p. 190).

TABLE III

LEATHERJACKET COUNTS

Treatment	1 (Control)	2 (Control)	3	4	5	6
Block I	92	66	19	29	16	25
II	60	46	35	10	11	5
III	46	81	17	22	16	9
IV	120	59	43	13	10	2
V	49	64	25	24	8	7
VI	134	60	52	20	28	11

The original analysis was actually made for $\sqrt{(x+\frac{1}{2})}$ for the *treated* plots only, the numbers on the control plots being used merely to indicate the degree of control by the treatments (see Table IV below).

TABLE IV

SUMMARY OF RESULTS

	1 2	3	4	5	6	S.E.	Sig. diff. ($P = .05$)
Mean $\sqrt{(x+\frac{1}{2})}$	....	5.58	4.43	3.84	3.03	0.407	1.23
Mean No./plot	73.1	31.8	19.7	14.8	9.8	....	....
o/o Control	....	56	73	80	87	....	....

A comprehensive analysis of variance including the control plot numbers is somewhat a matter of convenience when the control plot

[3] Beall gives a table in Reference 7 for values of $k = \lambda^2$ from 0 to 1. Since this table is for λ^{-1} Sinh^{-1} $[\lambda\sqrt{x}]$ and not λ^{-1} Sinh^{-1} $[\lambda\sqrt{(x+\frac{1}{2})}]$, I would recommend the empirical correction of replacing zero values of x by $\frac{1}{4}$ (cf. section ·4).

[4] See also the examples in Reference 7.

numbers differ considerably from the treated plot numbers. Here if it is desired to include control plot numbers in the analysis they might still reasonably be included in the square-root analysis, but the use of the more general variance law (6) would be safer.

It has been noted above that when biological populations change, the change is often proportional to the mean, implying changes independent of the mean on the logarithmic scale. However, suppose the fraction of area covered by a species of plant is the measurement; there is then a factor limiting the amount of growth, the fractional area never exceeding 1. In such situations I have found the transformation to the scale $\log \{x/(1-x)\}$ useful (cf. Reference 4, p. 163).

It is of interest to add to our list the well-known transformation used for a sample correlation coefficient r to make its distribution less skew and more stable in variance; viz., $\frac{1}{2} \log \{(1+r)/(1-r)\}$. Since the variance of a correlation coefficient is approximately $(1-\rho^2)^2/(n-1)$, where ρ is the true value of the coefficient and n the number of observations in the sample, we obtain this transformation from our equation (3) if we wish to make the variance independent of ρ. It is, of course, rare to have to analyze a set of correlation coefficients by analysis of variance, but if the problem arose the above transformation would be the appropriate one.

A more important problem that does frequently occur is the analysis of variance of a set of sample variances or standard deviations. A detailed discussion of this problem has been given elsewhere (Reference 6), and as I have already quoted the illustrative example once in this country,[5] I will not do so here, but merely note that the variance of a sample variance s^2 is proportional to $(\sigma^2)^2$, and hence the logarithmic transformation is suitable.

4. *The Inverse Sine or Angular Transformation.* The inverse sine square-root transformation

$$g(x) = \text{Sin}^{-1}\sqrt{x} \tag{7}$$

bears the same relation to estimated probabilities or proportions x with binomial variance $p(1-p)/n$, where n is the number of individuals in the sample, as the square-root transformation does to a Poisson variate. The approximately constant variance on the new scale is $821/n$, provided that the inverse sine, which denotes an angle, is measured in degrees. A table in this form is given in Fisher and Yates' *Tables* (Reference 17, p. 42; see also Reference 11). An alternative table in

[5] In a paper "Applications of Analysis of Variance," given at Princeton University on November 1, 1946.

radians was given in Reference 3, and on this scale the variance is $0.25/n$. In Tables V and VI below are quoted (Reference 4, p. 167) the data and analysis of results in one of a series of experiments for which this transformation was used in the routine analysis. It is not a particularly ideal "textbook" example, but is useful as an example of the rough evaluation of insecticides in contrast with detailed evaluations for which the probit transformation (see section 5) is more appropriate. The insecticides were here in the form of toxic sprays, and no exact dose for any insect is known.

TABLE V

NUMBER OF DEAD FLIES

Treatments	A	B	C	D	E	F	G*
1	24	(25)	17	17	18	23	1
2	25	(25)	15	17	25	25	1
3	24	(25)	12	17	24	23	1
4	21	(25)	20	22	16	23	10
5	25	(25)	21	13	22	23	4, 6

* Control (with one extra replication).

TABLE VI

SUMMARY OF RESULTS

	A	B	C	D	E	F	G	S.E.
o/o Kill	95	(100)	68	69	84	94	15	.. . .
$\sin^{-1}\sqrt{x}$ (radians)	1.41	(1.57)	0.98	0.99	1.22	1.34	0.37	0.082

In the above analysis no correction for "discontinuity" was used, since adding one-half to the observed numbers cannot consistently be carried through to the top end of the scale, near 100 percent kill. It was, however, pointed out in a footnote to my original discussion (Reference 4, pp. 167–168) that an empirical but fairly useful correction is simply to write $\frac{1}{4}$ wherever 0 occurs (and $n-\frac{1}{4}$, for n), and leave the other integers unchanged. This correction has a similar effect in "smoothing" the jumps due to the data consisting of whole numbers, the most violent jumps on the transformed scale being from 0 to 1 (or from $n-1$ to n).

In the theoretical discussion of Poisson and binomial variation by Cochran (Reference 14), already referred to in section 2, Cochran has pointed out (p. 346) that in an exact analysis of percentages the above empirical correction would become replaced by special adjustments, but he also notes that such an analysis would only apply to binomial data. It thus appears that the empirical correction I have suggested will remain useful in practical applications. For example, in the series of insecticide experiments referred to above, the mean variance was of the order of 0.03, as against $\frac{1}{4} \times 1/25 = 0.01$ for binomial variation, so that the assumption of exact binomial variability would certainly not have been tenable.

5. *The Probit Transformation.* For details of the probit transformation reference should be made to Bliss' original papers (References 9 and 10). This transformation, which converts the relative frequency with which a normal deviate y is exceeded into the corresponding value of y, is particularly useful when such a transformed quantity y is linearly dependent on another variable x, so that the transformation converts the functional relation between these two variables to a straight line. It is well known in toxicological investigations that this is often achieved if x denotes the logarithm of the dosage,[6] the relative frequency measured being the proportion of animals surviving at any particular dosage. However, the method is quite general and is often useful in other fields [see, for example, Reference 5].

The fact that the variance is not constant on the transformed scale implies that the observations must be weighted, and this is a disadvantage if analyses of variance involving more than one classification are required. But single classification analyses of variance are, of course, readily made. For example, for the data quoted in Table VII (see Reference 4, p. 165, and Reference 3, p. 188), the difference between the fitted regression lines for the two groups is readily tested in analysis of variance form by first working out the regression lines for each group separately in the usual way. This analysis for each line (see Reference 9) would be represented algebraically by the scheme:

$$\begin{array}{lllll}
Swx^2 & Swxy & Swy^2 & n \text{ d.f.} & \text{(a)} \\
(Swx)^2/Sw & (Swx)(Swy)/Sw & (Swy)^2/Sw & 1 \text{ d.f.} & \text{(b)} \\
\hline
Sw(x-\bar{x})^2 & Sw(x-\bar{x})(y-\bar{y}) & Sw(y-\bar{y})^2 & n-1 \text{ d.f.} & \text{(c)}
\end{array}$$

where w denotes the weight, x log-dosage, and y probit value, and S summation over the n observations. Fitted regression line:

$$Y - \bar{y} = b(x - \bar{x}), \text{ where } b = Sw(x-\bar{x})(y-\bar{y})/Sw(x-\bar{x})^2$$

and residual sum of squares

$$Sw(y-\bar{y})^2 - [Sw(x-\bar{x})(y-\bar{y})]^2/Sw(x-\bar{x})^2 \text{ with } n-2\text{d.f.}$$

[6] The transformation of the *independent variable* (e.g., to the logarithm or reciprocal) in regression problems is of course a common and valid procedure.

TABLE VII

FUMIGATION EXPERIMENT (24 HOURS EXPOSURE) ON THE BEDBUG

Dose	Adults		Nymphs	
(mg./liter)	Total	Dead	Total	Dead
7.83	20	2	10	0
11.76	25	3	9	5
17.20	24	6	5	3
19.00	23	6	11	10
20.90	24	19	7	6
23.20	22	8	10	5
24.60	13	8	18	17
28.00	9	9	27	24
29.80	28	21	4	3
32.00	25	19	3	3
36.90	15	14	17	17
40.20	17	17	15	15
44.90	8	8	20	20
Totals	253	140	156	128

If rows (a) are now added for each group, and a new row (b) obtained from the total sums $Swx = S_1wx + S_2wx$, etc., we shall obtain a new row (c) with $n_1 + n_2 - 1$ d.f., and a new residual sum of squares with $n_1 + n_2 - 2$ d.f. Subtracting the sum of the two residual sums of squares for the two groups with $(n_1 - 2) + (n_2 - 2)$ d.f., we have a term with 2 d.f. representing the difference in the two regression lines. If further we merely pool the two rows (c) to form a new row (c), we should eliminate separately the means of the two groups and obtain a residual sum of squares with $n_1 + n_2 - 3$ d.f. The difference between this sum of squares and the sum of the original two residual sums of squares for the two groups now has only 1 d.f. representing the difference in slopes of the two regression lines. The other d.f. in the 2 d.f. previously obtained represents a further difference in position of the two lines.

In this way the following analysis of variance (Table VIII) was obtained for the data of Table VII.

TABLE VIII

	d.f.	Sum of squares	Mean square
Adults v. Nymphs	2	26.226	13.113
{ Difference in position	1	26.216	.... }
{ Difference in slope	1	0.010	.... }
Residual	22	41.194	1.872
Total	24	67.420	....

The highly significant difference between the two groups is entirely arising from a simple parallel displacement of one line relative to the other as is evident from the separate regression equations, viz.:

$$\text{Adults:}\ Y = 4.985x - 1.563,$$
$$\text{Nymphs:}\ Y = 5.081x - 0.838.$$

The same situation was also found for a further experiment at $16\frac{1}{2}$ hours exposure (see Reference 2, Table IV).

A third transformation for percentages is noted in Fisher and Yates (Reference 17), the transformation log $\{p/(1-p)\}$. This transformation(mentioned in section 3) would, like the probit transformation, not give for the binomial variate a variance constant on the new scale, but might sometimes be useful for the other reasons. Thus probabilities often combine in a multiplicative way which might sometimes be more simply dealt with on this scale when corresponding estimated probabilities are available from data. Berkson (Reference 8) has even suggested that this transformation, which he calls the "logit" transformation, owing to its relation with the logistic function of population growth (cf. Reference 4, p. 163), is more useful than the probit transformation in bio-assay. The validity of such a claim must of course rest ultimately with experience, and evidence from other investigators on the relative value of these two transformations for particular types of data would be useful.

6. *Expected Normal Scores for Ranked Data.* It sometimes happens that the data presented to the statistician are not measurements, but a set of ranks; e.g., an investigation might have been made to determine the effect of manuring treatments on the *quality* of oranges and several buyers have been asked to grade different specimens, corresponding to different manuring treatments, in order of preference. This might have been repeated for several replications. Suppose there were r replications, b buyers and t treatments. Then it would be tempting

TABLE IX

SCHEMATIC ANALYSIS OF VARIANCE OF RANKS

	d.f.	S.S.	M.S.
Between treatments	$t-1$	x	x
Treatments × replications	$(t-1)(r-1)$	x	x
Treatments × buyers	$(t-1)(b-1)$	x	x
Treatments × buyers × replications	$(t-1)(b-1)(r-1)$	x	x
Total	$(t-1)rb$	$t(t^2-1)\,rb/12$	$t(t+1)/12$

to analyze these ranks in the usual analysis of variance way. The analysis would look as in Table IX.

The treatments term would be compared with the treatments × replications interaction; this test being based on the average opinion of the buyers. If, however, the treatments × buyers was significant compared with the second order interaction, it would indicate anomalous opinions among the buyers which would condition the conclusions from the first test.

With such data the variance is automatically stable, being determined by the ranks 1, 2, . . ., t for each order of preference given. There is no doubt that while, owing to the distribution of these ranks not being normal and the ranks having to some extent pre-determined scores, the above analysis is only approximate; it would at the same time often be useful. However, it is reasonable to assume that if the ranked data were replaced by expected normal scores, the validity of the analysis of variance would be somewhat improved, so that such a transformation might often be worth while. Fisher and Yates (Reference 17) give a table of such transformed scores, with corresponding sum of squares which would replace the quantity $t(t^2-1)/12$ in the above analysis.[7]

It has been suggested by various writers (see, for example, Reference 18) that even when measurements are available it may be safer to analyze by use of ranks. If so, it is permissible further to transform to these expected normal scores. It might, however, be remembered that if we discard the original measurements apart from their order we are throwing away the original scale and all quantitative transformations of it, one of which may well be relevant for estimating quantitative treatment effects and measuring interactions; such wholesale jettisoning should be avoided if possible.

7. *Scales Chosen for Additivity.* It was pointed out in section 1 that requirement (d), that the transformed scale should be one for which real effects are likely to be linear and additive, will not necessarily be consistent with requirement (a), and that the choice of scale will depend on the nature of our analysis of variance problem. In section 5 the probit transformation was an example of a scale not chosen in regard to (a).

A problem of scaling for *non-numerical* classified data on the basis of requirement (d) is discussed by Fisher (Reference 16, p. 285). Twelve samples of human blood tested with twelve different sera gave

[7] For an actual numerical example of a similar analysis, see Reference 12.

reactions denoted (in order of strength of reaction) by the symbols –, ?, *w*, (+) and +. The values 0 and 1 were assigned to the symbols – and + respectively and it was found that the values for the other symbols which maximized the ratio of the combined sum of squares for (i) the differences between the samples and (ii) the differences between the sera, when compared with the residual sum of squares, were 0.19 for ?, 0.58 for *w*, and 0.96 for (+). These values might thus be used as scores for the various reactions, on the argument that the interactions of samples and sera have been reduced to a minimum and the scale chosen thus made most nearly additive.

However, it will be appreciated by most practical statisticians that such detailed investigations on scale are not warranted on many data. Even for the data just referred to, Fisher notes that the scale represented by the above scores is not significantly different from the scale 0, $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$, 1; it can also be noticed from the data (Reference 16, Table 61.9) that one blood sample and one serum gave all *w* reactions, and while this might be regarded as evidence of the findings that the *w* reaction is somewhat isolated in "distance" from the others, it also underlines Fisher's comment that more extensive data are desirable for such an analysis.

Analogously to the above choice of scale for non-numerical data, it is possible to investigate transformations of scale for numerical data by which the ratio of the sum of squares for group differences to the residual sum of squares is maximized.[8] There are, however, as already stressed, theoretical complications in appraising significance in an ordinary analysis of variance if the scale is not selected for constant variance; this further complicates the present problem, since the variance at most will only be constant on one of the scales. If a statistician contemplates undertaking such an investigation, he should first be quite clear that he understands the basis and any limitations of the technique he is proposing to use and that its application to his own numerical material promises to be worth while.

REFERENCES

1. Bartlett, M. S. "Square-root Transformation in Analysis of Variance," *Journal of the Royal Statistical Society, Suppl.*, 3 (1936), 68.
2. Bartlett, M. S. "Some Notes on Insecticide Tests in the Laboratory and in the Field," *Journal of the Royal Statistical Society, Suppl.*, 3 (1936), 185.
3. Bartlett, M. S. "Subsampling for Attributes," *Journal of the Royal Statistical Society, Suppl.*, 4 (1937), 121.
4. Bartlett, M. S. "Some Examples of Statistical Methods of Research in Agriculture and Applied Biology," *Journal of the Royal Statistical Society, Suppl.*, 4 (1937), 137.

[8] Cf. J. W. Tukey's paper on "Vector Methods in Analysis of Variance," given at Princeton University on November 1, 1946.

5. Bartlett, M. S. "A Modified Probit Technique for Small Probabilities," *Journal of the Royal Statistical Society, Suppl.*, 7 (1946), 113.

6. Bartlett, M. S., and Kendall, D. G. "The Statistical Analysis of Variance-Heterogeneity and the Logarithmic Transformation," *Journal of the Royal Statistical Society, Suppl.*, 7 (1946), 128.

7. Beall, G. "The Transformation from Entomological Field Experiments So That the Analysis of Variance Becomes Applicable," *Biometrika*, 32 (1942), 243.

8. Berkson, J. "Application of the Logistic Function to Bio-Assay," *Journal of the American Statistical Association*, 39 (1944), 357.

9. Bliss, C. I. "The Calculation of the Dosage-Mortality Curve," *Annals of Applied Biology*, 22 (1935), 136.

10. Bliss, C. I. "The Comparison of Dosage-Mortality Data," *Annals of Applied Biology*, 22 (1935), 307.

11. Bliss, C. I. "The Analysis of Field Experimental Data Expressed in Percentages" (in Russian), *Plant Protection*, Leningrad, 1937, fasc. 12, 67.

12. Bliss, C. I., Anderson, E. O., and Marland, R. E. "A Technique for Testing Consumer Preferences, with Special Reference to the Constituents of Ice Cream," *Storrs Agriculture Experiment Station Bulletin*, 251, 1943 (Univ. of Connecticut).

13. Cochran, W. G. "Some Difficulties in the Statistical Analysis of Replicated Experiments," *Empire Journal of Experimental Agriculture*, 6 (1938), 157.

14. Cochran, W. G. "The Analysis of Variance When Experimental Errors Follow the Poisson or Binomial Laws," *Annals of Mathematical Statistics*, 11 (1940), 335.

15. Curtiss, J. H. "On Transformations Used in Analysis of Variance," *Annals of Mathematical Statistics*, 14 (1943), 10.

16. Fisher, R. A. *Statistical Methods for Research Workers.* Oliver and Boyd, Edinburgh, 9th ed. 1944.

17. Fisher, R. A., and Yates, F. *Statistical Tables for Biological, Agricultural and Medical Research.* Oliver and Boyd, Edinburgh. 1938.

18. Friedman, M. "The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance," *Journal of the American Statistical Association*, 32 (1937), 675.

APPENDIX

Summary of Transformations

Variance in terms of mean m	Transformation	Approximate variance on new scale	Relevant distribution
m	$\sqrt{x}$, (or $\sqrt{(x+\frac{1}{2})}$	0.25	Poisson
$\lambda^2 m$	for small integers)	$0.25\lambda^2$	Empirical
$2m^2/(n-1)$	$\log_e x$	$2/(n-1)$	Sample variances
$\lambda^2 m^2$	$\log_e x$, $\log_e (x+1)$ $\log_{10} x$, $\log_{10} (x+1)$	λ^2 $0.189\lambda^2$	Empirical
$m(1-m)/n$	$\text{Sin}^{-1} \sqrt{x}$, (radians) $\text{Sin}^{-1} \sqrt{x}$, (degrees)	$0.25/n$ $821/n$	Binomial
$m(1-m)/n$	Probit	Not constant*	
$m(1-m)/n$	$\log_e [x/(1-x)]$	$1/[mn(1-m)]$	
$\lambda^2 m^2 (1-m)^2$	$\log_e [x/(1-x)]$	λ^2	Empirical
$(1-m^2)^2/(n-1)$	$\frac{1}{2} \log_e [(1+x)/(1-x)]$	$1/(n-3)$	Sample correlations
$m+\lambda^2 m^2$	$\lambda^{-1} \text{Sinh}^{-1} [\lambda \sqrt{x}]$, or $\lambda^{-1} \text{Sinh}^{-1} [\lambda \sqrt{(x+\frac{1}{2})}]$	0.25	Negative binomial
$\mu^2(m+\lambda^2 m^2)$	for small integers	$0.25\mu^2$	Empirical
....	To expected normal scores	1 for large n*	Ranked data

* See Fisher and Yates (Reference 17) for exact values.

ANNUAL MEETING OF THE BIOMETRICS SECTION

The annual business meeting of the Biometrics Section was held December 29, 1946, in the Statler Hotel, Boston, Massachusetts, with President D. B. DeLury in the chair and 29 members present. Brief reports were heard on the reorganization of the American Statistical Association and on changes in the *Biometrics Bulletin* (now *Biometrics*). The report of the nominating committee, consisting of Churchill Eisenhart, Boyd Harshbarger (Chairman), J. A. Rigney and G. W. Snedecor, listed for Chairman: D. B. DeLury; Secretary: H. W. Norton; Section Committee: Geoffrey Beall, E. J. DeBeer, D. B. DeLury, D. J. Finney, H. W. Norton and J. W. Tukey; and Editorial Committee: R. L. Anderson, C. I. Bliss, W. G. Cochran, Gertrude M. Cox (Chairman), Churchill Eisenhart, H. W. Norton, G. W. Snedecor and C. P. Winsor. In the absence of other nominations, this slate was unanimously elected. Later, the Section Committee met and discussed means of broadening the Section's activities and membership.

A NOTE FROM THE EDITORIAL COMMITTEE

This is the first issue under our new format, which should improve the readability of articles. It is also the first issue under the shortened title of *Biometrics*. It is hoped that we may soon be able to expand into a Journal form. A slight backlog of articles is being accumulated, but not enough to put us on safe ground as yet. As a result of the expanding activities of the Biometrics Section, we have requested several authorities in associated fields to act as collaborating editors in securing and editing articles. To date we have received favorable replies from:

W. J. Dann, Duke University School of Medicine.
D. J. Finney, Lecturer in the Design and Analysis of Scientific Experiment, University of Oxford.
G. E. Dickerson, Regional Swine Breeding Laboratory, Bureau of Animal Industry, Ames, Iowa.
H. O. Halvorson, The Medical School, University of Minnesota.
C. M. Mottley, Chief Eastern Inland Fishery Investigations, Fish and Wildlife Service, Washington, D. C.
J. G. Osborne, Forest Service, Washington, D. C.

QUERIES

CORRECTION IN QUERY NO. 33, AUGUST, 1946

Leo A. Aroian has called attention to an oversight in the answer to this question. In *Biometrika*, Volume 33, pages 78–88, Merrington and Thompson have given "Tables of Percentage Points of the Inverted Beta (F) Distribution" which include 2.5 and 0.5 percent points. These points are therefore the desired 5% and 1% levels for the symmetrical test.

(43)

QUERY: It appeared to me that there was a typographical error on page 9 of the *Biometrics Bulletin* for February (Vol. 2, No. 1). In the third line from the bottom the average value for the mean square is given as $V + 12V_{ep} + 4V_r$. Should not the coefficient of V_r have read 48 instead of 4?

ANSWER: Yes.

S. Lee Crump

(44)

QUERY: The simple correlation between mothers and daughters for winter clutch size is + .1385. There were 237 mothers and 2407 daughters. In the calculations each daughter was paired against her mother. In interpreting the significance of r should we consider 236 degrees of freedom?

ANSWER: There seems to be no simple rule for cases such as this. The significance is certainly somewhat more assured than if the data were restricted to 237 mothers, each with only one daughter, but certainly far less than if there were 2407 daughters each from a different mother. The number of degrees of freedom which would give the proper test of significance is somewhere between the 235, which would be proper if each dam had only one daughter, and two less than the geometric mean of 237 and 2407, as can be seen from the following more precise way of getting at it.

Designate the observed correlation of + .1385 as x and let y be the correlation between the clutch size of the dam and the average clutch size of her k daughters. Then $y = x\sqrt{\dfrac{k}{1+(k-1)v}}$ where v is the intra-class correlation between daughters of the same dam. In the present query the size of v is not given, but from the nature of the material it is probably about the same size as x—perhaps a little larger and almost certainly larger than x^2. When k is variable its proper value in the above formula is its harmonic mean, which of course will be smaller than its arithmetic mean. Assuming that v has the same value as x in the present data, and using 9 for k, we get $+.29$ for y. Clearly y if computed directly would have 235 degrees of freedom and a value of $+.29$ would certainly be significant well beyond the .01 level. The significance of x (which seems to be the practical point the querist had in mind) will be the same as that of y.

That the ratio y/x depends not alone on the size of k but also on the size of v will show why no simple statement of the number of degrees of freedom proper for testing the significance of x can always be correct. The number of dams and the number of daughters and the size of x simply do not contain all of the information needed. The size of v is pertinent in that if v is very high, one daughter tells us nearly as much about what kind of daughters that dam will have as a large number of daughters would, while if v is very low two daughters tell nearly twice as much as one, four daughters tell nearly as much as two, etc. Another way to say it is to consider that x is the product of three links: record of dam to real value of dam, real value of dam to real value of daughter, and real value of daughter to record of daughter. Having many daughters per dam permits the variations in the last two links to cancel each other but has no such effect on any sources of variation which may have made low the correlation between record of dam and real value of that dam.

It may also be worth noting that, if the dams were themselves selected on clutch size, the observed r of $+.14$ is probably smaller than would have been found in a wholly unselected population. If such selection is thought to have occurred, regression of daughter on dam would be more appropriate than correlation.

JAY L. LUSH

(45)

QUERY: I was interested in seeing your "Question and Answer" column recently and decided to submit a problem that has been bothering me in connection with my work here in a government agency.

One of my assignments has been to determine the "average yearly rate of increase (or decrease)" in a series of annual data. Thus if my data are

1941	100
1942	120
1943	110
1944	140
1945	130,

I have always figured that from 1941 thru 1945 the total change has been +30 percent and, taking the fourth root of 1.30 I get an average annual rate of increase of 6.78 percent.

One of my fellow workers argues that I should average the individual yearly percents of change. In my illustration, that would be an average of +20%, −8.33%, +27.27% and −7.14% which gives 7.95 percent as the answer!

As a result of our argument, we consulted a statistician from another branch of the government, and he, using the same figures as above, worked out a "least squares exponential equation" and determined that the average yearly rate of increase was 7.025 percent!

May I ask a second question: In computing the first quartile from data in an array, should one use the value of the $\frac{N+1}{4}$ item or should one first compute the median and then determine the value of the $\frac{N+1}{2}$ item of the first half of the items?

ANSWER: As is usual with queries of this nature, the "right" procedure depends on the basic hypotheses associated with the origins of the data, as well as the purpose to which the desired "average" is to be put.

Method number one corresponds to finding the uniform growth rate, stated in terms of percent per year, such that a figure increasing uniformly with that growth rate would increase from 100 in 1941 to 130 in 1945. Stated algebraically, the curve $y = Ke^{\beta x}$ is fitted so that it will pass through the points (1941, 100) and (1945, 130). A growth rate established this way may be used as a descriptive figure, generally, in cases where the random fluctuations are small compared with the systematic growth, and where the departure from uniformity over the period in question is not too great. As an example, consider the growth of a baby in the first year to be $\alpha\%$. Then the "average monthly rate of increase" is $100\left(\sqrt[12]{1+\frac{\alpha}{100}}-1\right)\%$. This furnishes a common basis against which to compare the actual non-uniform monthly growth percentages for each particular month of the first year.

Method number three corresponds to the hypothesis that $y = Ke^{\beta x+\varepsilon}$, where the ε's are considered as independent values of a random variate; in other words, that the growth may be considered to be essentially uniform except for random errors, perhaps of measurement, or due to statistical perturbations. The other essential point here is that one wishes to predict y for given x values. If the above hypothesis is satisfied then the curve $y = Ke^{\beta x}$ fitted by least squares on the logarithm of y, by considering the regression of log y on x, gives a prediction which minimizes the variance of the residual of log y.

The second method corresponds to a hypothesis which is quite different, namely that the various values of $y(x+1)/y(x)$ for the successive values of x represent essentially independent observations on a random variate whose distribution does not depend on x. In this case, to predict $y(x+1)$ the formula $y(x+1) = \beta y(x)$ might be used, where β is determined from the "average rate" as determined by method number 2. Comparing with method three, we see that our hypothesis in number 2 is that the latest observation contains within it the best basis from which to project to the next, whereas in number 3 it is assumed that the deviation of the last observation from a fitted curve has no effect on the deviation which may occur in the next.

In answer to the second question, the first quartile is the $(k+1)$st observation when N is of the form $4k+1$; in other words, the first quartile is given by x_{Q_1} with $Q_1 = \frac{N+3}{4}$ if $\frac{N+3}{4}$ is an integer. If $\frac{N+3}{4}$ is not an integer one has a choice of procedures, depending on the form of distribution postulated. The usual procedure is to take the two observations corresponding to the integers next larger and next smaller than $\frac{N+3}{4}$, i.e., such that $n_1 < \frac{N+3}{4} < n_2$, and then average those two observations with weights depending on n_1, $\frac{N+3}{4}$, and n_2. For example one might compute

$$\frac{\dfrac{x_{n_1}}{\dfrac{N+3}{4} - n_1} + \dfrac{x_{n_2}}{n_2 - \dfrac{N+3}{4}}}{\dfrac{1}{\dfrac{N+3}{4} - n_1} + \dfrac{1}{n_2 - \dfrac{N+3}{4}}}$$

G. W. Brown

ABSTRACTS*

(34)

LOTKA, ALFRED J. (Metropolitan Life Insurance Company, New York). **Some Applications of the Life Curve.**

Principal fields in which the life curve is used are public health, insurance, population analysis, and the theory of industrial replacement.

Public health—The life table is a measure of the state of health of the population. It is not ideal because it takes no account of morbidity, except insofar as this finds expression in mortality. The expectation of life at birth is frequently used as a single-number index of health conditions. While it is not ideal, a high correlation has been found between the expectation of life at birth and the death rate in certain populations. Applied to the individual states of the United States in 1929–31, the correlation, using crude death rates, was $-.752 \pm .063$; using standardized rates, it was much higher, $-.992 \pm .002$.

The life table is also used for computing the years of life lost annually by deaths from individual diseases, and conversely the number of years that might be saved by the elimination of these causes of death. The effect of eliminating two or more causes jointly is greater than the sum of the separate effects.

Applications in population analysis—One is the construction of life tables, not for individuals but for families, that is, lines of male descent. We may ask, for instance, out of 1,000 newborn males, how many will have descendants in direct male line surviving after 1, 2, 3, etc., generations. A life table of this character was exhibited.

Population growth—Under certain conditions population growth approaches an exponential law. Ordinarily, the development of the formulas has used integral equations. A modification, using the raw data without introducing a continuous curve for the net fertility, was exhibited together with a tabular schedule of computation. In this method factorial moments enter in place of the ordinary moments which appear when the integral equations are used.

The intrinsic rate of natural increase in many European countries has been negative in past decades, while in the United States, until the time of the war, it was barely positive. If, after the present higher birth rates are past, fertility continues to decline as formerly, or if

* All papers read at the joint meetings of the Biometrics Section and the Institute of Mathematical Statistics held in conjunction with the 113th Annual Meeting of the American Association for the Advancement of Science in Boston, December 27–29, 1946.

a severe increase in the death rate should occur resulting from atomic bombing or biological warfare, a critical situation may arise. This draws attention to the unusual situation in which the human species finds itself. Man has been extraordinarily successful in fighting off other species. Consequently, the life struggle is today mainly a competition between man and man. In this competition the extraordinary skills and expedients which he has developed, instead of being an advantage to him, may actually prove his undoing.

(35)

DEEVEY, EDWARD S., Jr. (Osborn Zoological Laboratory, Yale University, and Woods Hole Oceanographic Institution). **Life Tables for Natural Populations of Animals.**

Materials for the construction of ecological life tables, in the form of (a) survivorship data for marked individuals, or (b) known age at death of large numbers of individuals, are available from several natural populations of animals. A third type of information, (c) the age structure of the natural population, is frequently obtainable from fisheries work. In no case have ecologists been able to observe directly both the age structure and the time-specific mortality rate for given age groups. However, if it is assumed that the population is at equilibrium, so that the actual age structure and the life table age structure are identical, information of the second and third types can be used in constructing life tables which are valid for particular ecological conditions, though their statistical foundation may not be rigid. Information of the first type can be used without qualification, when, as is usually the case, the animals have a sharply defined season of birth. When a cohort of such animals is born at a particular "instant," and its survival is followed until terminus, age-specific death rates are directly observed.

Survivorship curves have been prepared for several species of animals living under natural conditions. *Age at death* (b) has been observed for the Dall mountain sheep (Murie), the herring gull (Paynter, unpublished), and for a sessile rotifer (Edmondson). *Age structure* (c) is known (beyond certain ages) for the common tern (Austin), for many species of fish, of which the haddock (Russell) is an example, and for the female fin whale (Wheeler). *Survivorship of marked individuals* has been followed in the case of the song sparrow (Nice), the pheasant (Leopold), the snowshoe rabbit (Green and Evans), and a barnacle (Hatton).

When such l_x curves are compared, two important points emerge. (1) The survivorship of animals in nature sometimes, but by no means always, follows the J-shaped distribution (Pearl's "Type D") resulting from extremely heavy mortality at early ages. Diagonal l_x curves (Pearl's "Type B") are evidently frequent in natural as well as in laboratory populations. (2) The form of the l_x curve is strongly affected by its point of origin. Bird populations, for example, can be considered from the time the eggs are laid, from hatching, from fledging, or from breeding age, and correspondingly different life tables will result. It is not easy to say which one should be compared with the life table for a mammal or an invertebrate. The point of universal biological equivalence for animals is doubtless fertilization of the ovum, but in no case can the origin of a life table be taken at so early an age from existing data.

Life tables are extremely useful vehicles for comparative vital statistics, and ecologists are urged to make greater use of this method of presentation. Its advantages are illustrated by an analysis of Hatton's data for the survival of barnacles, which were allowed to settle on experimental panels under various conditions of exposure to surf and to intertidal desiccation, and followed throughout life. Taking expectation of life (e^0_x) as the criterion of survivorship, it can be shown that survival is an inverse function of population density. This analysis is based on a theory of two-dimensional crowding, in which geometric considerations make it possible to include both the radial growth and the initial density in evaluating the degree of crowding of sessile organisms on a panel.

(36)

DeLURY, D. B. (Virginia Polytechnic Institute). **The Analysis of Covariance.**

A single physiological experiment is discussed in detail, bringing out some features of an experimental situation which call for covariance methods. Numerical illustrations of the computations involved in simple and multiple covariance are given.

(37)

BLISS, C. I. (Connecticut Agricultural Experiment Station and Yale University). **Biological Measurement of the Depth Dose of X-rays of Lettuce Seedlings.**

The data of an experiment by Henshaw and Francis on the biological effectiveness of x-rays in a paraffin phantom have been reexamined. Lettuce seeds were exposed for five different periods at six levels in a paraffin phantom. The root growth of the irradiated seeds and of the negative controls was measured after 96 hours of germination. On the hypothesis that growth was proportional to the number of cells surviving treatment, the logarithm of root length was plotted against the length of exposure. The observations at the six different levels were fitted by simultaneous equations with straight lines which converged at zero time.

The agreement of the observations with the computed lines was tested by an analysis of variance. Their linearity and convergence at zero time agreed with the hypothesis that a single hit is sufficient to kill a cell. The observations were more variable at the levels nearer the tube, emphasizing the importance in experiments of this type of adjusting the length of exposure in successive levels to compensate for inequalities of scattering and absorption in paths of different lengths.

The biological effectiveness of the radiation decreased in the paraffin as the 2.45 power of the value expected in air. Since the expected intensity of direct radiation decreased approximately as the 4.85 power of the value in air, scattered radiation accounted for a large part of the total effectiveness of the rays in paraffin. The variation among the 17 replicates of the experiment of the log-slopes for the six different levels was examined by an analysis of variance. It disclosed a significant variability in respect to both average susceptibility and the rate at which the potency diminished in the phantom.

(38)

YOUDEN, W. J. (Boyce Thompson Institute for Plant Research, Inc., Yonkers 3, N. Y.). **On the Question of Duplicate Analyses.**

Analyses conducted on samples of the same size yield a mean value and a standard deviation, but no information on any constant error in the analysis. If observations $y_1, y_2, \ldots, y_n$ are made on samples of weight $x_1, x_2, \ldots, x_n$, where the x's form a graded series, and a straight line, $y = a + bx$, is fitted by least squares, then $\underline{a}$ is an estimate of any constant error and $\bar{y}$ and $b\bar{x}$ are two independent estimates of the y associated with $\bar{x}$. If $\underline{a}$ is significantly different from zero then $b\bar{x}$ is to be preferred to $\bar{y}$ even though it has a larger variance. When the x's are of the form $1w, 2w, \ldots, nw$, the variance of $b\bar{x}$ is $\frac{\sigma^2}{n} \times 3\frac{(n+1)}{n-1}$ as compared with the variance $\frac{\sigma^2}{n}$ for $\bar{y}$.

(39)

KNUDSEN, LILA F. and JACK M. CURTIS. (Food and Drug Administration). **The Use of the Angular Transformation in Biological Assays.**

A simplified method is given for evaluating biological assays having quantal responses. The probit, logit, and angular transformations are compared. Use of the angular transformation is advocated to make the weights used proportional to the number of animals on that particular dose. It is shown that if equal numbers of animals are used on each dose of standard and unknown of a two-dose assay with a constant ratio of doses $\left(i = \log \frac{\text{high dose}}{\text{low dose}} = \text{constant}\right)$, a graph for determining potency and a nomograph for estimating the error of the assay can be constructed for use in the laboratory.

A comparison of results calculated by the probit method with those calculated by the angular transformation method shows that the two methods give practically the same results but that the amount of time required to make the calculations on a two-dose assay by the two methods is very different. Probit method calculations using the exact weights require one and a half or two hours; the angular transformation method requires eight to ten minutes.

NEWS AND NOTES

This issue of NEWS AND NOTES was prepared by the acting editor, while EDITOR COX cavorted beneath the spreading palms at Honolulu (and also taught some classes in Experimental Design at the Pineapple Research Institute). She left Raleigh on December 19 and is expected back around April 1. Meanwhile, we received such inspiring messages as, "The moon last night was the most romantic moon I've ever seen. . . . Am going native. The messenger boy is supposed to bring me a fresh hibiscus each morning." Also, such a typical occurrence as this was reported, "I talked the full hour today, then dismissed the class. A group of men from Hawaiian Pineapple stayed and we had an hour and a half discussion on some of their work." Shades of past performances in design classes! Incidentally, for the benefit of those readers who may have stopped off in Hawaii during the recent world conflagration, Miss Cox stayed at the Moana Hotel.

R. N. JEFFERSON, who was with the Experiment Station at V.P.I., Blacksburg, Virginia, before entering the army, has taken a position as Assistant Professor of Entomology of the University of California and Assistant Entomologist in the Agricultural Experiment Station. He is working on insects which attack ornamental plants. He writes, ". . . and Southern California is a pretty nice place to live." He was married December 24 to the Canadian girl whom he met at Iowa State College. . . . OSCAR KEMPTHORNE, formerly of the Rothamsted Experimental Station, has joined the staff of the Statistical Laboratory at Iowa State College. . . . PAUL DENSEN has been appointed Chief of the Division of Medical Research Statistics, Bureau of Medicine and Surgery, Veterans Administration, Washington, D. C. The position of Assistant Professor of Preventive Medicine and Public Health at Vanderbilt University, vacated by Dr. Densen, has been filled by MARGARET MARTIN, formerly Assistant Professor of Statistics at the University of Minnesota. . . . G. J. FISCHER, Instituto Fitotecnico, Estanzuela, Uruguay, has requested a picture of those attending the 1946 summer school of the Institute of Statistics. He also reports that a recent guest at the Instituto was F. G. BRIEGER, who gave a lecture on the Indian corn of Montevideo. . . . E. C. FIELLER writes that his permanent address is now the Department of Scientific and Industrial Research, National Physical Laboratory, Teddington, Middlesex. . . . ABD EL MONEIM ASHEUR is Director of the Animal Production Department, Faculty of Agriculture at Farouk First University. His mailing address is 18 Rue Fouad First, Alexandria, Egypt. . . . W. A. TIMMERMAN, formerly of Johannesburg, South Africa, is now located in the Insurance Building, 907 15th St., N.W., Washington, D. C. . . . E. L. WILLETT, formerly of the University of Hawaii, can be reached at the American Scientific Breeding Institute, 134 N. LaSalle Street, Chicago. . . . W. G. MATHENY has transferred from the Experiment Station at Fort Hays, Kansas, to the Department of Psychology, University of Maryland.

A letter has just been received from H. FAIRFIELD SMITH, now back at the Rubber Research Institute of Malaya, P. O. Box 150, Kuala Lumpur, Federated Malay States. It has been rumored that he was a prisoner of war during the Japanese control of Malaya, although no news of his war experiences was given in his letter. Almost all of his reprints were lost, and I am sure he would appreciate receiving any reprints which the readers of *Biometrics* might have available. Fairfield Smith's work on discriminant functions as applied to plant selection has been widely used in recent years by animal as well as plant breeders. A brief article by him on an approximation to the number of degrees of freedom in a composite variance is now being extended to both analysis of variance problems and to an approximate solution of the classic Behrens-Fisher problem. . . . The nature of the Fairfield Smith approximation to the solution of the Behrens-Fisher problem was discussed by M. S. BARTLETT at a University of North Carolina seminar on January 29. This seminar preceded a farewell banquet given for Professor Bartlett who has returned to Cambridge University after a four-months stay with the Institute of Statistics.

Material for *Biometrics* should be addressed to the Chairman of the Editorial Committee, Institute of Statistics, North Carolina State College, Raleigh, N. C.; and material for Queries should go to ''Queries,'' Statistical Laboratory, Iowa State College, Ames, Iowa, or to any member of the committee.